JAN 15 1934
EDUCATIONAL LIBRARY
UNIV. OF MICH.

The JOURNAL *of* EXPERIMENTAL EDUCATION

Volume II | SEPTEMBER, 1933 | Number 1

METHODS OF INSTRUCTION

CONTENTS

Introduction: GOOD, CARTER V. ..... *i*

The Effect of Educational Motion Pictures upon the Retention of Informational Learning: HANSEN, JOHN ELMORE 1

An Evaluation of an Arithmetic "Crutch": BROWNELL, WILLIAM A. 5

"Intelligence" and "Perseveration" Related to School Achievement: ROGERS, K. H. 35

The Effect of Praise, Reproof, and Exercise upon Muscular Steadiness: EATON, MERRILL T. 44

Measuring Attitude toward the High School: REMMERS, H. H., BRANDENBURG, G. C., and GILLESPIE, F. H. 60

The Measurement of Attitudes toward Current Political and Economic Problems among Junior and Senior High School Pupils: WHITTAKER, MILO L. 65

$5.00 A YEAR | PUBLISHED QUARTERLY | $1.50 A COPY

EDWARDS BROTHERS, INC.
LITHOPRINTERS AND PUBLISHERS
ANN ARBOR, MICHIGAN

Application filed for entry as second class matter. Printed in U. S. A.

JOURNAL OF EXPERIMENTAL EDUCATION

EDITORIAL BOARD

A. S. Barr, Chairman, Professor of Education, University of Wisconsin, Madison, Wisconsin.

Carter V. Good, Professor of Education, University of Cincinnati, Cincinnati, Ohio. Editorially responsible for materials on supervision and psychology of learning and teaching.

Henry Harap, Associate Professor of Education, Western Reserve University, Cleveland, Ohio. Editorially responsible for materials on experimental studies of curriculum construction.

Walter S. Monroe, Professor of Education, University of Illinois, Urbana, Illinois. Editorially responsible for materials on measurements, statistics, and methods of experimental research.

George D. Stoddard, Director, Child Welfare Research Station, State University of Iowa, Iowa City, Iowa. Editorially responsible for materials on child welfare, guidance, and development.

CONTRIBUTING EDITORS

Harry J. Baker, Director, Psychological Clinic, Detroit Public Schools, Detroit, Michigan.

W. E. Blatz, Nursery School Division, University of Toronto, Toronto 5, Canada.

William F. Book, Head, Department of Psychology and Philosophy, Indiana University, Bloomington, Indiana

Fowler D. Brooks, Head, Departments of Education and Psychology, DePauw University, Greencastle, Indiana.

William A. Brownell, Professor of Educational Psychology, Duke University, Durham, North Carolina.

Leo J. Brueckner, Professor of Education, University of Minnesota, Minneapolis, Minnesota.

Herbert B. Bruner, Professor of Education, Teachers College, Columbia University, New York City.

Barbara S. Burks, Psychologist, Institute of Child Welfare, University of California, Berkeley, California.

Otis W. Caldwell, Professor of Education, Teachers College, Columbia University, New York City.

Ellsworth Collings, Dean, College of Education, University of Oklahoma, Norman, Oklahoma.

Philip W. L. Cox, Professor of Secondary Education, New York University, New York City.

Edgar A. Doll, Director of Research, Training School, Vineland, New Jersey.

Harl R. Douglass, Professor of Education, University of Minnesota, Minneapolis, Minnesota.

Jack W. Dunlap, Associate Professor of Education, Fordham University, New York City.

Paul H. Furfey, Professor of Psychology, Catholic University of America, Washington, D. C.

Florence L. Goodenough, Professor, Institute of Child Welfare, University of Minnesota, Minneapolis, Minn.

M. E. Herriott, assistant director Division of Psychology and Educational Research, Public Schools, Los Angeles, California.

Karl J. Holzinger, Professor of Education, University of Chicago, Chicago, Illinois.

L. Thomas Hopkins, Curriculum Specialist, Lincoln School of Teachers College, 425 West 123rd Street, New York City.

C. L. Huffaker, Professor of Education, Bureau of Educational Research, University of Oregon, Eugene, Ore.

Kai Jensen, Assistant Professor of Education, University of Wisconsin, Madison, Wisconsin.

Harold E. Jones, Director of Research, Institute of Child Welfare, University of California, Berkeley, California.

Edward A. Lincoln, Assistant Professor of Education, Graduate School of Education, Harvard University, Cambridge, Massachusetts.

E. F. Lindquist, Professor of Education, State University of Iowa, Iowa City, Iowa

Lois Hayden Meek, Professor of Education, Director, Child Development Institute, Teachers College, Columbia University, New York City.

C. W. Odell, Associate Professor of Education, University of Illinois, Urbana, Illinois.

Willard C. Olson, Director of Research in Child Development, University of Michigan, Ann Arbor, Mich.

W. E. Peik, Associate Professor of Education, University of Minnesota, Minneapolis, Minnesota.

S. L. Pressey, Professor of Educational Psychology, Ohio State University, Columbus, Ohio.

W. H. Pyle, Professor of Educational Psychology, Detroit Teachers College, Detroit, Michigan.

Paul T. Rankin, Supervising Director, Board of Education, Detroit, Michigan.

Clarence E. Ragsdale, Assistant Professor of Education, University of Wisconsin, Madison, Wisconsin.

H. H. Remmers, Director, Division of Educational Reference, Professor of Education and Psychology, Purdue University, Lafayette, Indiana.

G. M. Ruch, Professor of Education, University of California, Berkeley, California.

Earl U. Rugg, Head, Department of Education, Colorado State Teachers College, Greeley, Colorado.

Peter Sandiford, Professor of Educational Psychology, Director, Educational Research, University of Toronto, Toronto, Canada.

Douglas E. Scates, Director, School Research, Cincinnati Public Schools, Cincinnati, Ohio.

Raleigh Schorling, Professor of Education, Supervisor, Directed Teaching and Instruction in University High School, University of Michigan, Ann Arbor, Michigan.

Mandel Sherman, Associate Professor of Education, University of Chicago, Chicago, Illinois.

Helen Thompson, Research Associate, Yale University, New Haven, Connecticut.

Herbert A. Toops, Professor of Psychology, Statistics and College Personnel, Ohio State University, Columbus, Ohio.

T. L. Torgerson, Assistant Professor of Education, University of Wisconsin, Madison, Wisconsin.

M. R. Trabue, Executive Secretary of Committee on Individual Diagnosis and Training, Institute for Employment Stabilization Research, University of Minnesota, Minneapolis, Minnesota.

R. W. Tyler, Professor of Education, Ohio State University, Columbus, Ohio.

Douglas Waples, Graduate Library School, University of Chicago, Chicago, Illinois.

Beth L. Wellman, Research Associate Professor, Child Welfare Research Station, State University of Iowa, Iowa City, Iowa.

The JOURNAL *of*
EXPERIMENTAL EDUCATION

Volume II September, 1933 Number 1

INTRODUCTION

It is a far cry in psychology to the establishment of the first psychological laboratory in America and to the founding of the American Journal of Psychology, both accomplishments of G. Stanley Hall. Various schools of psychology have developed, which include the introspective and structural groups, behaviorism, Gestalt, psychoanalysis, and purposivism. The early beginnings in educational psychology, as represented by Stanley Hall's introduction of experimental psychology into America and his initiation of the child-study movement, have extended to a considerable number of educational problems and fields. In addition to publications which discuss educational psychology and learning in somewhat general terms, there are specialized psychological treatments of such problems as mental hygiene, study, elementary-school subjects, secondary-school subjects, arithmetic, algebra, athletics, play, music, typewriting, skill, handwriting, reading, spelling, character, personality, social institutions, language, adolescence, childhood, adult learning, etc.

The current issue of the Journal of Experimental Education is devoted to studies of the psychology of learning and teaching. The articles in this number represent techniques of research and fields which merit brief introductory comment.

The two studies by Hansen and by Brownell are evaluations of teaching method. Hansen, using a parallel-group technique of experimentation, has investigated a problem in the important field of visual education. Brownell has evaluated painstakingly and minutely the effect of a device employed in the teaching of fractions in arithmetic. Not the least significant part of this report is the detailed account of the various steps and leads followed in the attempt to solve the problem.

Rogers through elaborate statistical calculations has analyzed "factors" related to school achievement.

Eaton reports on experimental study in the field of motivation of learning.

Two studies deal with pupil attitudes, a problem of increasing concern in the modern school. Remmers et al. present a scale for measuring attitude toward high school. Whittaker reports a detailed investigation of the political and economic opinions of secondary-school pupils.

Carter V. Good

Carter V. Good

THE EFFECT OF EDUCATIONAL MOTION PICTURES UPON THE RETENTION OF INFORMATIONAL LEARNING

by
John Elmore Hansen[1]
University of Wisconsin

PURPOSE OF STUDY

The purpose of the investigation reported herein is to determine the contribution of educational motion pictures to the retention of informational learning.

PREVIOUS STUDIES

During the past fifteen years numerous studies have been made to determine the contribution of educational motion pictures to learning, noteworthy among which are the studies of Weber[2], Freeman[3], Wood and Freeman[4], Knowlton and Tilton[5], and the Middlesex and Historical Association studies[6] in Great Britain. These studies have shown, without exception, that the educational motion pictures, when properly used, are effective aids to learning. The Knowlton and Tilton study in New Haven and the Historical Association study in England appear to be the only studies, however, in which much attention was given to the influence of motion pictures upon the retention of what was learned. Knowlton and Tilton measured the retention of historical knowledge over intervals varying from approximately three to seven months. The results of this study indicate that the experimental group had learned more, forgotten more, and remembered more than the control group.[7] A summarized statement from The Film in National Life concerning the British Historical Association study says: "The film leaves a remarkably permanent impression. Deferred tests were given in both inquiries up to seven months after the film had been shown, and produced surprisingly good results, particularly where the subject of the film had been related to the school curriculum and had been followed up by regular lessons."[8] With this rather limited information it would seem that further evidence is desirable to test the permanency of learning made with the aid of motion pictures.

SUBJECTS

The subjects employed in this investigation consisted of approximately two hundred fifty pupils, including the entire student body of the Middleton High School and the pupils of the seventh, eighth, and ninth grades of the Mt. Horeb Public Schools. The data presented herein are for eighty-two matched pairs of subjects. The chief facts concerning these subjects are given in Table I. The experimental and control groups employed in this investigation were eighty-two

1. The writer wishes to express his appreciation of the kind cooperation and guidance extended by Dr. A. S. Barr, Professor of Education at the University of Wisconsin in the conduct of this experiment. Thanks are also due to Principal L. W. Amborn of the Middleton High School and to Principal Harley J. Powell of the Mt. Horeb Public Schools, in whose schools the study was made.

2. Joseph J. Weber, Visual Aids in Education (Chicago: The Educational Screen, Inc., 1930).
3. Frank N. Freeman, Visual Education (Chicago: The University of Chicago Press, 1924).
4. Ben D. Wood and Frank N. Freeman, Motion Pictures in the Classroom (Boston: Houghton Mifflin Company, 1929).
5. Daniel C. Knowlton and J. Warren Tilton, Motion Pictures in History Teaching (New Haven, Connecticut: Yale University Press, 1929).
6. Commission on Educational and Cultural Films, The Film in National Life (London: George Allen and Unwin Ltd., 1932).
7. Daniel C. Knowlton and J. Warren Tilton, Motion Pictures in History Teaching (New Haven: Yale University Press, 1929), p. 57.
8. Commission on Educational and Cultural Films, The Film in National Life (London: George Allen and Unwin Ltd., 1932), p. 67.

TABLE I
AGE, SEX, AND INTELLIGENCE OF THE SUBJECTS EMPLOYED IN THIS INVESTIGATION

Group	No. of Pupils	No. of Boys	No. of Girls	M. A. Average	I. Q. Average	Initial Test Average
Experimental	82	38	44	181.3	101.6	38.4
Control	82	36	46	180.8	101.9	38.0

pairs of subjects matched according to mental age, intelligence quotient, and previous knowledge of the topic to be studied as measured by an initial test of information. The mean M. A. of the Experimental Group was 181.3 and of the Control Group, 180.8; the mean I. Q. of the Experimental Group was 101.6 and of the Control Group, 101.9. The Experimental Group was slightly higher in M. A. and slightly lower in I. Q. The number of girls in the Experimental Group was 44 and in the Control Group, 46; the number of boys was 38 and 36 respectively. Since psychologists[1] are almost unanimous in agreeing that there is little difference in the ability of the sexes to learn, it would seem that little would be gained in matching according to sex.

MATERIALS OF INSTRUCTION

The topic studied by the two groups was that of the making of pottery. This topic was one with which none of the pupils were likely to be familiar and one about which they had little previous information. The motion picture film used was the Eastman Teaching Films, Inc., production, "Tableware." Following the opening scene which depicts primitive pottery making, it shows in detail the various steps involved in the manufacture of earthenware in a modern pottery factory. It is a 16-mm. film, one reel in length, and required fourteen minutes for presentation. The reading matter consisted of a two-thousand word written exposition which described in detail all of the steps in the various processes pictured in the film. Every possible effort was made to make the written exposition and the film as nearly comparable as it is possible to make two such different kinds of presentation.

THE TEST

The test employed in the investigation consisted of three parts, totalling 102 items: Part I was made up of twenty-seven statements, each of which was to be completed by indicating one of five given terms as the correct one; Part II consisted of eleven questions, each of which had five statements concerning the same operation or object, any or all of which might be correct or wrong in which each pupil was asked to mark each of the statements indicating whether it was true or false; Part III consisted of twenty true-false statements which were to be marked either true or false.

The test was given three times. It was given first ten days previous to the presentation of the film and the reading material. This test served two purposes: first, as an aid in equating the groups, and, second, as a means of determining the initial knowledge of the topic already possessed by the pupils. The lapse of a ten-day period before administering the test a second time also served to minimize any practice effect. The second application of the test was made immediately after the pupils had received their instruction in the topic to be studied before they had any opportunity to forget what they had learned. The third test was given three weeks and four days, in the case of the Middleton School, and three weeks and two days, in the case of the Mt. Horeb Schools, after the second test to measure the effects of the two modes of presentation upon the retention of what had been learned as shown by the second test.

The reliability of the test was determined by the split-test method for the scores on the second and the third applications of the test. The coefficient for the second

1. Rudolph Pintner, Intelligence Testing (New York: Henry Holt and Company, 1931), pp. 496-501.

application of the test (T_2) is .765, and for the third application of the test (T_3), .815. The intercorrelation coefficient between the results on the first application (T_1) and the second application (T_2) is .359, and between the second (T_2) and the third (T_3), .939. Since the results gained on T_1 are an indication of the information already possessed by the subjects, which had been gained through random reading and observation, and since the additional information measured in T_2 and T_3 had been obtained under the same closely controlled conditions, it is to be expected that a lower correlation would exist between T_1 and T_2 than between T_2 and T_3. These reliability coefficients taken as a whole are sufficiently high for the purposes of this study.

No attempt has been made to determine the validity of the test. The items of the test were, however, carefully selected to cover the information given in the written exposition and the film. In making this statement the impossibility of constructing a verbal test which will adequately measure all of the contributions made through the means of a pictorial presentation is fully realized. It would appear that verbal tests favor learning gained through verbal presentation.

PRESENTATION OF SUBJECT MATTER

The presentation of the subject matter to the pupils was as follows:

Experimental Group

5 minutes - Orientation period
14 minutes - Viewing of film
10 minutes - Oral discussion based upon study outline in hands of pupils
14 minutes - Studying written exposition

43 minutes - Total length of period

Control Group

5 minutes - Orientation period
14 minutes - Reading of written exposition
10 minutes - Oral discussion based upon study outline in hands of pupils
14 minutes - Studying written exposition the second time, using study outline as a guide

43 minutes - Total length of period

It will be noted that the only difference in the treatment of the two groups was the substitution of a fourteen-minute film presentation in the experimental set-up for fourteen minutes of reading in the control set-up.

The total control and experimental groups were each taught in three separate sections as follows:

Middleton High School

1. 9th - 12th Grades, inclusive
   Experimental Class 42 pupils
   Control Class 42 pupils

Mt. Horeb Public Schools

2. 9th Grade
   Experimental Class 23 pupils
   Control Class 23 pupils

3. 7th and 8th Grades
   Experimental Class 17 pupils
   Control Class 17 pupils

The subject matter was presented to all of the group sections by the same instructor and the tests were administered by the same individual. This insured, as nearly as possible, that the pupils in the various sections received the same treatment.

RESULTS

The mean scores of the eighty-two pupils in the Control Classes and the eighty-two pupils in the Experimental Classes on the three successive tests are presented in Table II. The mean scores for the Control Group were: for T_1, 38.0; for T_2, 66.1; and for T_3, 60.9. The mean scores of the Experimental Group were: for T_1, 38.4; for T_2, 70.0; and for T_3, 65.2. This represented a gain between T_1 and T_2 of 28.1 points, and a net gain between T_1 and T_3 of 22.8 for the Control Group; and corresponding gains of 31.6 and 26.8 respectively for the Experimental Group. A comparison of these

TABLE II
MEAN SCORES AND STANDARD DEVIATIONS FOR THE EXPERIMENTAL AND CONTROL GROUPS ON TESTS I, II, AND III; THE GAINS BETWEEN TESTS; AND THE DIFFERENCES IN GAINS BETWEEN GROUPS

	X GROUP			C GROUP			
Test	(M)	S.D.		(M)	S.D.		M_x-M_c
I	38.4	S.D.(dis.) S.D.(av.)	11.86 1.31	38.0	S.D.(dis.) S.D.(av.)	12.37 1.36	.4
II	70.0	S.D.(dis.) S.D.(av.)	10.34 1.14	66.1	S.D.(dis.) S.D.(av.)	11.85 1.31	3.9
III	65.2	S.D.(dis.) S.D.(av.)	11.06 1.22	60.9	S.D.(dis.) S.D.(av.)	10.17 1.12	4.3
II-I	31.6	S.D.(dis.) S.D.(av.)	12.98 1.43	28.1	S.D.(dis.) S.D.(av.)	13.80 1.52	3.5
III-I	26.8	S.D.(dis.) S.D.(av.)	13.26 1.46	22.8	S.D.(dis.) S.D.(av.)	12.69 1.40	4.0

results shows that the Experimental Group made an average gain of 3.5 points more between T_1 and T_2 than did the Control Group, and that when tested after an elapse of nearly a month the net average gain between T_1 and T_3 is 4.0 points more for the Experimental Group than for the Control Group.

This difference between the Experimental and Control Group gains from T_1 to T_2 is .31 of the standard deviation of the combined Experimental and Control T_2 scores[1]; the difference between the Experimental and Control gains from T_1 to T_3 is .37 of the standard deviation.[2] In terms of per cent the differences in gains in favor of the Experimental Group are 12.1% and 17.5% respectively. The Standard Error[3] of the first difference is 1.615 and of the second, 1.62, giving critical ratios of 2.17 and 2.46 respectively. These differences are believed to be significant.

Two conclusions may be drawn from a study of the above results: (1) The educational motion picture is an aid to learning, especially in the acquisition of information such as portrayed in the film used in this study. This bears out the results of other studies mentioned earlier in this paper. (2) It seems that the knowledge gained through the means of pictorial presentation is fully as permanent in its nature as that gained through reading and oral presentation. Apparently the impressions made by combinations of pictures and words are more lasting than those produced by words alone. It is possible that the association of verbal forms with the appropriate visual images enabled the pupils in the Experimental Group to gain clearer and more definite notions of objects and their movements; thus the materials becoming more meaningful are learned more readily and retained longer than less meaningful material.

1. $S.D._{T_2(x+c)}$, 10.89

2. $S.D._{T_3(x+c)}$, 11.30

3. Standard Error of the difference in gains between groups of matched pairs has been computed as equal to

$$\sqrt{(S.D._{M_x})^2 + (S.D._{M_c})^2 - 2rM_xM_c}$$

M_x, mean Experimental gain
M_c, mean Control gain
r, coefficient of correlation between individual X gains and C gains.

AN EVALUATION OF AN ARITHMETIC "CRUTCH"

by
William A. Brownell
Duke University

PURPOSE OF THE STUDY

The major purpose of this article[1] is to report an investigation concerning the effects of instruction in the use of an arithmetic "crutch" in adding proper fractions. Data will be presented and analyzed with respect to three phases of the problem: first, the degree to which the "crutch" tended to persist in the arithmetic behavior of the children who had been taught to employ it; second, the extent to which use of the "crutch" involved children in uneconomical and even harmful procedures in adding proper fractions; and third, the relative amounts of ability to add fractions exhibited by children who had been taught the "crutch" and by comparable children who had not been taught this device.

The report is divided into two parts which are dissimilar in intent and in content and which are designed to accomplish two unlike purposes. The first part is an account of the steps in the investigation. It is little concerned with the exact data collected, but it is much concerned with the way in which the problem was almost accidentally discovered and with the manner in which through the application of appropriate research techniques the problem, none too clear at times, was finally made to yield to a solution. This part of the report is frankly intended to serve two ends. The first is to furnish a realistic example of the way in which much research is done: with periods of alternate confusion and enlightment, with the commission of many errors of judgment and the correction of those errors, and with the following of faulty leads and the attainment of fortunate successes. (Such a description of research method, it will be noted, bears little resemblance to the descriptions to be found in manuals and in many theoretical discussions of research.) The second end served by this part of the report is to supply evidence that not all research is spiritless grubbing and boring drudgery, but that on occasion it affords all the excitement and all the thrills enjoyed by a detective at work on his favorite type of crime.

Part II, by contrast, has little about it to make it emotionally attractive. Here will be presented a large share of the data which were collected in the course of the investigation, together with the evaluations and comparisons by means of which these data were studied to discover the relation of the use of the "crutch" to the development of ability to add proper fractions. While the statistical treatment which will be described is relatively simple, it is exceedingly detailed, and the reader who finds little to interest him in detail may be well advised not to read Part II at all. This part of the report is intended to supply the factual basis for Part I and to illustrate to the reader who is curious about such procedures certain useful methods of analysis which are not commonly employed.

Before beginning with Part I or skipping to Part II the reader will need certain information regarding the setting of the investigation, the techniques employed and the measures collected. This information is supplied in the following two sections.

SOURCE OF THE DATA

Data for this report were collected from 248 fifth-grade children[2] who constituted

1. The writer is deeply obligated to Dr. B. R. Buckingham, of Harvard University, and to Dr. Charles H. Judd, of the University of Chicago, both of whom read the original manuscript and offered much helpful criticism.
2. Data are actually reported for 245 children only, since the records of three pupils were not susceptible to accurate interpretation.

the total enrollment in seven schools,[1] three of them in Nashville, Tennessee, and four of them in adjacent Davidson County, Tennessee.

The test used was a form of the Brueckner Diagnostic Test in the Addition of Fractions[2] which was abridged and modified for the special purposes of this study. As revised, the test below consists of fifteen examples which classify into three major types. In the first five examples the common denominator occurs in both the fractions to be added; in examples 6-9 the common denominator occurs in only one of the fractions; and in examples 10-15 the common denominator must be found.

TEST USED IN THE INVESTIGATION

ADD:

(1) $\frac{1}{5}$	(2) $\frac{1}{8}$	(3) $\frac{2}{7}$
$\frac{2}{5}$	$\frac{5}{8}$	$\frac{5}{7}$
(4) $\frac{4}{5}$	(5) $\frac{4}{6}$	(6) $\frac{1}{2}$
$\frac{2}{5}$	$\frac{4}{6}$	$\frac{1}{4}$
(7) $\frac{1}{2}$	(8) $\frac{2}{3}$	(9) $\frac{4}{8}$
$\frac{2}{8}$	$\frac{3}{6}$	$\frac{3}{4}$
(10) $\frac{3}{8}$	(11) $\frac{1}{5}$	(12) $\frac{2}{4}$
$\frac{1}{12}$	$\frac{2}{4}$	$\frac{7}{11}$
(13) $\frac{3}{7}$	(14) $\frac{3}{6}$	(15) $\frac{7}{9}$
$\frac{3}{4}$	$\frac{4}{5}$	$\frac{5}{6}$

THE "CRUTCH"

The "crutch" which is the object of investigation in this study may be illustrated as it was employed in performing the addition required by the first example:

(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\frac{1}{5}$	=	$\frac{1}{5}$	$\frac{x}{x}$	$\frac{1}{1}$	=	$\frac{1}{5}$
$\frac{2}{5}$	=	$\frac{2}{5}$	$\frac{x}{x}$	$\frac{1}{1}$	=	$\frac{2}{5}$
						$\frac{3}{5}$

Its cumbersomeness and artificiality are at once apparent: the re-writing of the unchanged fractions (3), the double appearance of the multiplication sign after each fraction (4), the multiplication by $\frac{1}{1}$ (5), and so on. It is true that the multipliers in (5) varied somewhat from paper to paper and that there were other minor differences, but the general scheme with its unusual and perhaps unwarranted degree of elaborateness remained surprisingly constant.

Part I

EVOLUTION OF A RESEARCH PROJECT

Discovery of the "crutch." The use of the "crutch" was not discovered until some weeks after the fraction test was administered in the seven schools during the last week of January and the first week of February, 1930. As a matter of fact, the investigation proper had been originally designed to isolate and classify children's errors and difficulties in adding proper fractions as these were directly observable or indirectly procurable through individual interviews. The test papers, on which the children recorded the work which they described orally, were collected only on the chance that they might afford additional information of value with respect to a complete picture of children's methods in adding fractions.

1. The writer herewith acknowledges his obligation to Acting Superintendent J. J. Keyes, of the Nashville city schools, and to Superintendent W. C. Dodson, of the Davidson County schools, as well as to the principals, teachers, and pupils of the seven schools involved in the study.

2. L. J. Brueckner, Brueckner Diagnostic Tests in Fractions, Minneapolis, Minnesota: The Educational Test Bureau.

In the course of the tabulation of data from the test papers the "crutch" was noted in the work of the children in two of the schools (hereafter designated as schools 1 and 2). Its consistent appearance on the papers from these schools and its complete absence on the papers from the other schools (schools 3 to 7) was indeed striking and at once raised a number of questions.

Reason for interest. Further study of the test papers in schools 1 and 2 revealed the fact that the "crutch" had been employed with all three types of example in the test. As it was used with the second type of example it regularly had the appearance (example 6):

$$\frac{1}{2} = \frac{1}{2} \times \frac{2}{2} = \frac{2}{4}$$

$$\frac{1}{4} = \frac{1}{4} \times \frac{1}{1} = \frac{1}{4}$$

$$\frac{3}{4}$$

Its form in connection with the third type of example was as follows (example 10):

$$\frac{3}{8} = \frac{3}{8} \times \frac{3}{3} = \frac{9}{24}$$

$$\frac{1}{12} = \frac{1}{12} \times \frac{2}{2} = \frac{2}{24}$$

$$\frac{11}{24}$$

Now the possible usefulness of such an elaborate device in solving the last six problems is understandable, but its value in solving the easier examples of the first and second types is by no means obvious. It was apparent that the pupils in these two schools had been taught the complicated written device at the time they were first introduced to the addition of fractions and that they had been trained to employ it through all their later work in adding fractions.

Two facts of considerable interest stood out in clear relief. First, here was a case of deliberate instruction in the use of a "crutch", a practice which on theoretical grounds is freely condemned by many writers on Educational Psychology and on the teaching of arithmetic.[1] Second, the situation in which the "crutch" was being employed was such as to permit at least a partial evaluation of its effectiveness and its harmfulness through the agency of quantitative research techniques.

The literature on the place of "crutches" in instruction is highly controversial. It is likewise highly opinionated, for experimental data on the value and dangers of "crutches" are exceedingly rare. All too frequently the worth of a given instructional device is determined by some type of a priori classification. If the device seems to violate certain psychological principles which are dear to him, the opponent of "crutches" is apt to call it a "crutch"; if it appears to be useful, he will call it something else. On the other hand, the device is a "crutch" to the exponent of "crutches" if it appears to facilitate learning, and it is something else if it interferes with learning. In both cases, the decision is based upon preconceived notions, and in neither case are the merits of the device satisfactorily evaluated by labelling it or by refusing to label it a "crutch." It ought to be possible to test the usefulness of instructional devices by studying the degree to which they advance or retard sound learning. Fortunately in the conditions which surrounded the use of the fraction form found in schools 1 and 2 an opportunity was afforded for just such an analysis of its influence upon learning. In large measure, then, this report may well be taken as an example of an attempt to substitute quantitative data for opinion, an effort to determine the value of an instructional device, not by attaching a

1. Some may be unwilling to accept the form described above as an example of a true "crutch." Unlike most "crutches" this form does represent an advanced instead of an immature procedure, a finished instead of a crude way of doing something. It had its origin, not in the naive thinking of pupils approaching the study of a new arithmetical process, but in the mathematical logic of that process and in the thinking of the teacher who had mastered that logic.
Nevertheless, the form is here regarded as a "crutch." The justification for so classifying it lies in the fact that the device was given the children at the beginning of their study of fractions as a temporary aid, as a means of explaining to them the nature of the process, with the full expectation that eventually this "crutch" would be abandoned for other more direct procedures. If to the reader this reason seems insufficient to justify calling the form a "crutch", he may substitute in his thinking some such term as "device" for the recurring "crutch" without in any serious way altering the validity of the comparisons which will be made.

name to it, but by measuring the results of its use. While the form will be referred to as a "crutch" this practice is not intended in any way, favorably or otherwise, to prejudge the case; it will be called a "crutch" because of its artificiality and its purpose, and it may still be called a "crutch" after the facts have been surveyed with regard to the reality of its contribution to learning.

(a) Persistence of the "crutch"

Results of the first test in school 1. The fraction test was given the 35 children in grade V of school 1. Their papers and the records of their oral reports showed that these children fell into three distinct classes with regard to their use of the "crutch." Nine of the 35 children did not use the "crutch" at all. Thirteen others used the "crutch" ten times, with the four examples of the second type (C.D. given in one fraction) and with the six examples of the third type (C.D. to be found). The remaining thirteen used the "crutch" in all fifteen examples quite irrespective, therefore, of the three types of examples in the test.

These facts suggested what might be called a process of selective elimination, that is, a series of stages in growth in which the elaborate form was progressively abandoned as the cumbersomeness of that form became increasingly apparent. All the children had received the same instruction from teacher and textbook and had been equally trained, it is to be supposed, in the use of the "crutch" as a means of adding fractions. But now, some weeks later, some of the children apparently had completely freed themselves from it: they appeared to have grown beyond the point where it contributed anything vital to their work, even with comparatively difficult examples. Certain others still needed it or preferred to use it, even with the simplest examples. Still others occupied an intermediate position between the two extreme groups. These last children had reached a stage in development where they seemed to differentiate between types of examples, with some of which they employed the form and with others of which they no longer needed it.

It was this variable elimination of the "crutch" which first attracted attention. There must have been some reason why these children differed so markedly in the degree to which they had sloughed off use of the complicated form in favor of more direct procedures. The hypothesis of selective elimination even if valid required some explanation itself: Why had some of the children entirely discarded the "crutch" while others were completely bound to it? If the extent to which the "crutch" was still employed, some weeks after instruction in its use, did actually correspond to stages in development, how long would it be before those who were most dependent upon it would give it up? In a word, what factors were responsible for the persistence of the "crutch"?

An answer to these questions was first sought in the qualities and characteristics of the children themselves. As complete measures as possible were secured on various aspects of general development and of arithmetical ability. As measures of general development the following were taken: C.A., from the school records, and I.Q. and M.A. from the National Intelligence Test which was given the children for this specific purpose. As measures of arithmetic ability two scores on the fraction test were already available, the number of examples correctly computed and correctly reduced and the number of examples correctly computed but not reduced. Measures of problem-saving ability were obtained by taking scores on the problem-solving test in the N.I.T. General arithmetical ability, it was thought, was best indicated by teachers' marks. The school records in the case of most of the children yielded arithmetic marks for the two preceding terms.

It was proposed to compare the distributions of the three groups of children (those who used the "crutch" 15, 10, and 0 times on the test) for these different measures on the chance that significant relationships might be revealed. Thus, if for example, the children in the 15-group were found to average 130 months in M.A., and those in the 10-group to average 140, and those in the 0-group to average 150, one might suspect that the tendency to abandon the "crutch" was related to the attainment

of a certain level of mental maturity.

Complete measures were to be had for only 31 of the original 35 children in the grade.[1] These separate measures, being in themselves essentially meaningless, were next subjected to various methods of statistical analysis (the nature of which, not here reported, will be found in Part II). For the purposes of Part I it will serve merely to report the results of these analyses. In general, they were negative: no evidence could be found of a connection, much less of a close causal connection, between the persistence of the "crutch" on the one hand and any of the measures here studied on the other hand. These findings were more than a little disconcerting. The methods of analysis had been adopted with full confidence in their efficacy and with the expectation that they would isolate the cause for the persistence and the elimination of the "crutch", in so far as causes may be inferred at all from close statistical agreement.

And yet, while the data seemed to prove conclusively that persistence of the "crutch" could not be accounted for in terms of any of the available measures, the hope that the theory of selective elimination might be established would not down. An examination of the means of the various developmental and arithmatical measures for the groups gave some basis for this hope. On the average the 10-group did better on the fraction test than did the 15-group and the 0-group. When scores were based on completely correct answers all except one of the twelve children in the 10-group equalled or surpassed the mean score of the 0-group (by a margin of three examples or more), and eight of the twelve surpassed the mean of the 15-group. When credit was allowed for unreduced answers, the 10-group were still considerably superior to the 0-group, but not quite as successful as were the 15-group. As determined from teachers' marks, the arithmetic ability of the 10-group was markedly better than that of the other two groups, all except three of the twelve 10-children averaging higher than the mean for the 0- and the 15-groups. They likewise showed some advantage over the other groups in average M.A. and in average I.Q. All of these facts taken together, it was felt, might mean that the 10-children, who showed some discrimination in their use of the "crutch", were actually at the most desirable stage of their development.[2] Those who had discarded the form altogether (the 0-group) may have done so prematurely; they may have progressed too rapidly. Likewise those who still slavishly used the "crutch", employing it mechanically even with the simplest examples, might represent cases of "arrested development."

But all this was based on hope and might easily constitute a case of wishful thinking. The facts regarding the elimination of the "crutch" were still to be established, to say nothing of the theory of selective elimination. Furthermore, there remained the data from school 2 which might confirm, refute, or merely modify the interpretations placed on the data for school 1. The next step then was to examine the test data and the various other measures for school 2.

Results of the first test in school 2. Unlike school 1 which has the semester system school 2 is organized on the quarter basis, with three terms of approximately three months each. Each grade thus comprises three groups of children. The lowest of these groups, that is, the one beginning the work of the grade, is known as the C-section, the two higher groups being designated in order by the letters B and A. The subject of fractions is introduced in the C-section of grade V. Accordingly, when the fraction test was given in this school for the first time in the first week of February, 1930, to the C- and B-sections, the children in the C-sections had completed their study of fractions about a month before, and the children in the B-sections, about four months earlier. At the time of the testing and interviewing no attention was paid as to whether the children came from the C-section or from the B-section except to make note of the fact on the blanks when the tests were begun. Forty-five children were tested.[3]

1. See Table I, Part II. The loss in cases was as follows: one from the nine in the 0-group, one from the thirteen in the 10-group, and two from the thirteen in the 15-group.
2. Data collected in another connection also revealed the fact that the 10-group were much less prone than were the 15- and the 0-groups to utilize faulty and indirect procedures in securing their answers.
3. See Table IV, Part II. Scores on the problem-solving test in the N.I.T. are not given. This is the only difference as compared with Table I for school 1.

When the data had been tabulated it was discovered that nineteen of the children had made use of the "crutch" in all fifteen of the test examples and that the remaining twenty-six had not used it once. There was, then, in this school no intermediate group such as was found in school 1, who employed the form with certain examples but not with others. Furthermore, all nineteen children who consistently used the form belonged to the C-section while twenty-three of the twenty-six who did not use it at all belonged to the B-section. To the nineteen C-section children who had just concluded their study of the addition of fractions the "crutch" was either a necessary procedure or was one which was used habitually if needlessly. On the other hand, the twenty-three children in section B, being now four months removed from their instruction in the addition of fractions, no longer needed or employed the "crutch."

But thus far, three cases have been omitted from the discussion. These three (Cases 131, 142, and 147) were enrolled in the C-section, yet behaved like pupils in the B-section in that they did not once use the form in taking the test. Here seemed to be some important exceptions to the remarkably consistent data furnished by the other 42 children in the grade. The uniqueness of their performance led to a study of their school records. There it was learned that two of them (Pupils 131 and 147) had failed of promotion to the B-section at the end of the preceding quarter, were repeating the work of the C-section, and had therefore just been taught for the second time how to add fractions. Instead, therefore, of constituting exceptions these two cases proved to be excellent examples of the theory which had been advanced: to the effect that children, once they had gotten what they could from the form, tended to discard it.[1] The four months elapsing since their introduction to fractions had served the same purpose for them as it had for their fellow-pupils who had been promoted to the B-section.

The data for the two sections of school 2 could thus be made to agree with respect to the tendency for children to retain or to abandon use of the "crutch." Those pupils in the B-section who had had time to grow away from the form had completely done so; the C-section pupils, by hypothesis, were still too close to their instruction in its use to be able to dispense with it, though they might be expected to do so after a reasonable length of time. But the abandonment of the form in the B-section could, of course, have been due to other causes than the one of lapse in time for instruction,--causes associated, for example, with certain qualities or characteristics of their general development or of their general arithmetic ability. Comparison of the means for the two sections on those traits for which measures were obtained showed the B-section to have made distinctly higher scores on the fraction test, a fact which was to have been expected from their relative school classification; but more importantly these comparisons showed the B-pupils to have had marked advantage in brightness (I.Q.) and in mental maturity (M.A.). Had it not been for the check supplied by the data in school 1, where no relation was found between intellectual factors and abandonment of the "crutch", the conclusion might well have been advanced from the data for school 2 that the extent to which the form was used was directly conditioned by the intelligence of the pupil.

Results of re-testing in school 1. While such a conclusion was untenable in the light of the different conditions in the two schools, still certain other inconsistencies made it equally dangerous to place all the emphasis on distance in time from instruction in the use of the "crutch" as the chief factor to account for its use or its elimination. This explanation fitted well enough the facts in school 2, but it did not at all fit the facts in school 1. In school 2 the time between initial instruction in adding fractions (and hence in the use of the "crutch") and the administration of the fraction tests differed for the B- and C-sections, and this time difference was almost exactly reflected in the relative degree to which the children in the two sections employed that form. But in school 1 the time between instruction and test was

1. It was, however, impossible to discover any explanation for the failure of Pupil 142 to use the "crutch" in the test. Accordingly in the comparisons later made Pupil 142 was retained in the C-section (Table IV) whereas Pupils 131 and 147, for the reasons given above, were omitted from the C-section.

uniform for all pupils, and yet there remained marked differences in the degree to which they were dependent on that device.

By the time that the data had been analyzed far enough to disclose the lack of agreement in the results for the two schools about three months had passed since the first testing and interviewing in school 1. In other words the fifth-grade children in this school were about four and a half months distant in time from their first instruction in the addition of fractions. So far as time relations between instruction and testing were concerned, then, the situation at that time (May) in school 1 was roughly similar to that in section B of school 2 when the test was given there originally. It was decided to give the fraction test a second time to the children in school 1. The expectation was that these children would show the same effect of time upon their use of the "crutch" as was manifested in the case of school 2. Other things being equal, all the children in school 1, it was thought, should on the re-test give evidence of having completely discarded the form.

But apparently "other things" were not "equal"; the re-testing failed utterly to produce the anticipated results. Six of the eight children in the 0-group on Test 1 used the "crutch" ten times on Test 2. All twelve of the 10-group on Test 1 used the "crutch" ten times or more on Test 2. None of the eleven children in the original 15-group had entirely freed themselves of the "crutch" in the four months' interval, only four of them using the form at all less commonly on Test 2.

Instead of clearing up the situation, therefore, the re-test data in school 1 further complicated it. It had seemed reasonable to believe that the extensive use of the "crutch" on Test 1 by the children in school 1 was due to their closeness to instruction in its use and to believe further that lapse in time would bring about a change. Lapse in time did bring about a change, but a change in the wrong direction, if the re-test data were to be accepted as valid. The interval of four and a half months from initial instruction in the use of the "crutch" had not only not led to its abandonment; it had fastened the "crutch" even more firmly upon its users. The disagreement in results for schools 1 and 2 had not been removed as anticipated; it had been increased.

Results of re-tests in school 2. It was impossible to drop the investigation at this point. The diverse sets of conditions in schools 1 and 2 simply could not both be equally true. It seemed scarcely reasonable that in school 2 the passing of a certain amount of time should witness a general growth toward independence from the "crutch" while in school 1 the passing of a longer period of time should witness no such change, or rather on the contrary should reveal a reversion of its use on the part of children who earlier appeared to have abandoned it. The need for further study was clearly indicated.

One explanation for the discrepancy in the two schools lay in the possibility that the data for school 2 with respect to use of the "crutch" might have been improperly interpreted. The C-children had been found to be completely dependent upon the "crutch", and the B-children to be completely independent of it. The assumption was made that the children who were in the B-section at the time of the test had, three months before, been fully as dependent on the "crutch" as were those who were then actually in the C-section. The assumption was further made that the C-section children three months later would be as free of the "crutch" as were the children actually in the B-section at the time of the test.

Assumptions of this kind are not infrequently made in studies of the type here reported, but there is always the danger that in a particular instance the assumption may be invalid. Fortunately it was not necessary in this study to leave in doubt the validity of the assumption. The 22 children in the C-section at the time of Test 1 were still available for re-testing three months later. All of them had for about one month been enrolled in Section B under a new teacher.

On the re-test not one of these 22 children, who on Test 1 had used the "crutch" consistently, made any use of that form.

The general direction of the change wrought during the lapse of time since Test 1 was not unexpected in view of the results from the first test in the two sections of this school. But the uniformity of the

change, the fact that it had taken place for all of the children, was unexpected.

And yet here again the cause of the elimination of the "crutch" need not have been merely time, with the opportunity it afforded for the discovery and adoption of simpler procedures in adding fractions. The cause might just as conceivably have been the influence of the new teacher. Accordingly she was informed of the nature of the investigation and was questioned concerning her observations of the children's use of the form. She had noticed, she said, that a few children (not nearly all of them, as was true at the time of Test 1) on first coming into Section B tended to employ the device. By far the majority of them, however, apparently had dispensed with it. The few who still relied upon it were easily led to give it up for new methods. Moreover, once learned, the new methods rapidly and thoroughly superseded the "crutch." Again she emphasized the fact that she encountered no instructional difficulty: such children as still employed the form seemed to be doing so largely in a mechanical fashion and took readily to suggested improvement.

Next an interview was held with the teacher who had given all the children in grade V, both the B- and the C-sections, their first training in adding fractions. The first explanations, she said, as well as all subsequent practice in adding fractions were in terms of the form. The children were required to show the procedure in all their written work, and solutions were regarded as wrong if the form did not appear on the paper or on blackboard. No exceptions were allowed. When she was told of its disappearance from the fraction habits of her pupils later in section B, she expressed no surprise. On the contrary, she stated that very frequently during their last weeks in the C-section numbers of children omitted the form in their work and that it was difficult to maintain regard for the rule for its use. They seemed to discover on their own initiative that this form no longer contributed to their success in adding fractions, and in spite of her active opposition they seemed to prefer their new self-evolved procedures.

Further testing in school 1. All of the data for school 2 were now internally consistent. But so were most of the data for school 1 internally consistent, and the stories told regarding the persistence and elimination of the "crutch" were diametrically opposed to each other. In school 2 the "crutch" was generally discarded when the children no longer needed it; in school 1 the "crutch" not only was not discarded, but it seemed to become more rather than less necessary as time went on. In fact the retest data in the latter school clearly justified the question as to whether the children in this school might ever later be found adding fractions without the device.

As one means of finding an answer to this question the fraction test was given to 32 children in grade VI of this school, all of whom had attended grade V in that school the year before and had been taught the "crutch" even as had the pupils who at the time of the investigation were in this grade. These sixth-grade children were, then, a year and several months removed from their instruction in the use of the "crutch" and were roughly comparable to the fifth-grade children in the same sense that the children in the B- and the C-sections of school 2 were comparable, except, of course, in school 1 a period of more than a year instead of three months separated the two groups.

Not one of the 32 children in grade VI used the "crutch" in the fraction test.

The observations of the sixth-grade teacher relative to the children's use of the form agreed exactly with those of the teacher of the B-section in school 2. At the beginning of the year, that is, a few weeks after leaving grade V in which the form was required, a small number of the children, she noticed, employed the form. The few were induced with a minimum of effort on her part and on theirs to substitute other more economical methods. Most of the children, she found, had already adopted such methods, apparently of their own accord, for they used them before she supplied them, and she knew that in grade V these methods were neither suggested nor permitted.

The conversation with the sixth-grade teacher together with the results of the testing in her grade tended to reduce somewhat the differences between schools 1 and 2. These new data from grade VI showed

that the general course of development away from the "crutch" might actually prevail in school 1 as it seemed to in school 2, in spite of the evidence to the contrary in the re-test data for grade V.

Subsequent correspondence with the teacher of grade V in school.[1] The one set of data, those for the re-tests in grade V of school 1, were the only ones which were now out of line with a consistent account of the experience of these children first in using the later in laying aside the "crutch." Realization of this fact two months after the re-testing, when the tabulation and interpretation of data had reached this point, led to a critical consideration of the conditions under which the re-tests had been given in this grade. These had been administered in May, the day before the term examinations in this school. In the midst of the complications which naturally arose from the unexpected introduction of these re-tests into an already full school program certain peculiarities and odd happenings during the testing, which under other circumstances would have been promptly noted, might easily, it was thought, have escaped the notice of the examiner. It was, therefore, not until this later date that one of these happenings was recalled and that its possible relation to the data was appreciated. Then the examiner remembered a statement which was made by the teacher to the pupils just as the test papers were being distributed: "Now, children, don't forget what I have told you about working these examples."

The school had already been closed for the summer for several weeks, and it was necessary to communicate with the teacher by mail. The following letter was sent to her:

Dear _____:

I am sure you will recall the two occasions on which I did some work in common fractions with the children of your grade.

I am writing now to secure from you certain information which may be of considerable importance in interpreting one very interesting thing we found in the work of your pupils. What this is, I cannot now tell you, for you might be influenced thereby in answering the question I will ask you. I can, however, assure you that it is nothing to the discredit of yourself or of your pupils.

Just as the children began to take the test in fractions for the second time (in May) I remember your saying to them, "Now, children, don't forget what I told you about working these examples." Perhaps the words are not exact, but the above conveys their general meaning. Now, what I want to know is, what was it that you were telling them to do? You apparently had something definite in mind, and you knew that the children would understand to what you referred.

Won't you please write me a note on the reverse of this letter and mail it to me at once in the self-addressed and stamped envelope I am enclosing? Just tell me what you meant by your statement. If you can't recall, then tell me that. You may take my word for the fact that I am not asking an idle question, and that I would not be bothering you after your many courtesies if it were not important. Your answer cannot possibly be used against you, even should I be inclined so to use it.

Very truly yours,
etc.

Three days later the following note was received from the teacher:

Dear Dr. Brownell:

In regard to your letter referring to the last test you gave my children, I shall be glad to tell you what I meant when I said, "Now, children, keep in mind what I have told you about working these examples."

What I had reference to was this--for them to show you exactly how they made the fractions similar by multiplying both terms by the same number. Still another thing I had in mind was for them to notice those fractions that were already similar and to simply work them as they were.

I hope that I have made this clear to you. If I have not, please notify me, and I shall be glad to clear this up in some way.

I should like to know just how the children came out in the last test, or rather both of the tests.

Sincerely yours,
etc.

The reader's attention is directed in particular to the second paragraph in the

teacher's letter, in which the children were being warned to "show exactly how they made the fractions similar by multiplying both terms by the same number." The teacher's direction could have meant but one thing to these children who had been required in their daily work with fractions to employ the "crutch": they were being told to use the "crutch" on the test. And this they did, not because they had not outgrown the need of it, nor because they were not bright enough or mentally mature enough to get along without it, but because they were accustomed to doing as they were told to do.

Conclusion to (a). The teacher's letter constituted the final crucial bit of evidence in the case: it accounted for the one exception which had blocked an earlier answer to the question which was being investigated. A curious circumstance which had little connection with the "crutch" as such had prevented the children on the re-test from showing the freedom from the form which it is probable they had developed; and as a consequence this one curious circumstance had postponed the end of the inquiry and had made necessary many excursions and side-trips which, but for this circumstance, would never have been made.

But now the data with respect to the persistence of the "crutch" were all in, and an answer had been obtained to the first question proposed for study. In the case of the children investigated one may say with assurance that the "crutch" had not become fixed as a permanent habit; rather that it tended to drop out of their fraction habits when other simpler procedures had been found. On this one point the device meets one test which is raised by the opponents of such elaborate instructional forms; it meets this one test, but it does not thereby prove anything regarding its value or its contribution to sound economical learning.

(b) Types of faulty procedures in adding proper fractions associated with use of the "crutch"

A second kind of objection. But those who oppose "crutches" would properly argue that their objections to these devices are several in number. They might concede the validity and adequacy of the evidence adduced regarding the persistence of the particular device here under investigation. They might grant that this "crutch" did not tend to get itself permanently and ineradicably into the behavior of its users, and still they might insist that the form should not be taught. They could argue, for example, that the "crutch" would certainly confuse rather than clarify the process of adding fractions; that it would block learning by encouraging children to remain at a lower level of efficiency and maturity than they would attain without it. Thus, this "crutch" might involve children in related roundabout and erroneous methods of thinking, and in turn these methods, employed to excess, might then interfere with the development of skill and understanding in adding fractions. Specifically, use of the "crutch" might lead to habitual failure to reduce answers, to long or incorrect procedures in securing common denominators, and so on. If the device tended to produce these effects in children, they would maintain, it certainly should not be taught them.

Such an argument cannot lightly be brushed aside as irrelevant and immaterial. Neither, in the instance of the particular "crutch" here being studied, could the argument be disregarded as inappropriate. On the contrary, that the "crutch" might actually be open to the objection seemed more than possible in view of the elaborateness of the form. There was, clearly, just one thing to do, namely, to test the soundness of the objection by resorting again to objective, quantitative evaluation.

Nature of the data. Fortunately the right sort of data were available. As has been previously explained, the object of the investigation of which the study here reported was a part was to secure a list of pupils' errors and difficulties in adding proper fractions. These errors and difficulties were collected by means of conferences which were held individually with all of the 245 children in schools 1-7 at the time they were given the fraction test.[1]

1. These interviews required nearly 70 hours of time. They could have been made only through the assistance of others. This assistance was provided by four graduate students enrolled in the writer's course, "Research in Arithmetic." It is a pleasure to give the credit which is due these four: Mrs. Flora Gillentine, Miss Virginia Howard, Miss Myra Hodges, and Mr. J. G. Garrett.

The resulting list of "faulty procedures" or "faults" as they will hereafter be designated originally comprised a total of 55.[1] Six major categories were recognized, these being "Faults in finding common denominators" (of which, for example, there were ten specific types and a miscellaneous group which occurred only rarely), "Faults in operating on denominators", "Faults in operating on numerators", "Faults in reducing fractions to lowest terms", "Faults in obtaining whole numbers in reducing improper fractions to mixed numbers", and a group of three "Miscellaneous faults" which applied to procedures relating to the example as a whole. It was these data, carefully classified, which, it was thought, provided the best basis for evaluating the second objection to the "crutch."

Complications. The theory here was that if use of the "crutch" did really interfere with learning, this effect should be discernible in the degree to which children taught to use the "crutch" also made use of the faulty procedures. Thus, if the "crutch" children used a greater variety of faulty procedures, or used them more generally, or used them oftener, then the weight of the argument would be against the "crutch." This line of thinking suggested the method of attacking the problem, which was to compare the occurrence of faults in the work of the 80 children in the two "crutch" schools with their occurrence in the work of the 165 children in the five "non-crutch" schools.

These comparisons, which at the outset appeared to be simple in nature and few in number, became increasingly complicated and numerous. As had been said, it had been decided to study the two groups of children with regard to their comparative use of faulty procedures. But the phrase "comparative use" proved soon to be much more easily repeated than interpreted. For example, how was "comparative use" to be measured? The first measure which came to mind, "number of appearances", was found to be wholly unsatisfactory, for it became clear that it made much difference how and where the fault appeared. Thus, a frequency of 10 meant one thing when it represented the number of times a given fault was employed by a single pupil and it meant another thing when it stood for the total use by eight or ten pupils.

The distinction to which attention is called in the last sentence is one which has more than statistical significance; it has much significance for instruction when its ramifications are understood. When each child in a class of 45 fails one or more times to reduce an improper fraction on a test, making this error a total of 75 times, the teacher recognizes the need of remedial instruction for the class as a whole. When, however, the total of 75 failures to reduce is confined to five or ten of the pupils, the teacher has quite a different task; now she proceeds through individual instruction.

The comparisons to be made in this study had, therefore, to take account both of the number of times faulty procedures were used (frequency) and of the number of children who used them (universality). The influence of the "crutch" on learning could not properly be evaluated by considering only one of these two measures of "comparative occurrence."

A second group of difficulties arose in connection with determining the faulty procedures, use of which was to constitute the basis for evaluation. If the comparisons were restricted only to specific faults, then something might be lost. For instance, the "crutch" children might differ markedly from the "non-crutch" children, not so much in their use of certain faulty procedures, as in the proportionate number who used faulty procedures regardless of type, or in the number of times they used all faulty procedures combined, or in the variety of faults to which they were subject. Steps had to be taken to assure that measures were obtained at all these points.

Nor were the complications yet ended. There remained the question of the trustworthiness of whatever differences were found. The simple process of subtracting the number of uses for the one group from the number of uses for the other group did of course yield a difference, but this difference itself demanded some study. After all, the "non-crutch" children outnumbered

1. The list is omitted from this report for two reasons. In the first place, it would require too much space. In the second place, the purpose of the report does not require detailed descriptions of all the faults. Such information will be provided only for those faults which for one reason or another need explanation.

the "crutch" children more than two to one, and on this account comparisons could not be directly made as to frequency or universality of use unless appropriate corrections were made. Furthermore, the "non-crutch" children represented enrollment in five schools; the "crutch" children in two schools. Differences in faults as between the two groups of children could not be ascribed to instruction or non-instruction in the "crutch" until due consideration had been given to the possible influence of this last named factor. In other words, differences which were not reliable differences contained little of worth for the purposes of evaluation.

Use of faulty procedures. A beginning was made by noting the relative number of times the "crutch" and the "non-crutch" children employed faulty procedures. As a first step the children in the two groups were separately arranged in order according to the total number of faults which appeared on their papers. From this array frequency distributions were made (Part II, Table V) to show the number of children in the two groups who had made use of faulty procedures certain numbers of times. The frequencies were then transmuted to per cents in order to make allowance for the difference in size of the groups. Finally the two per cent distributions were plotted in curves in order to permit a direct comparison of the "crutch" and the "non-crutch" children regarding their use of faulty procedures. Figure 1 contains the two curves,

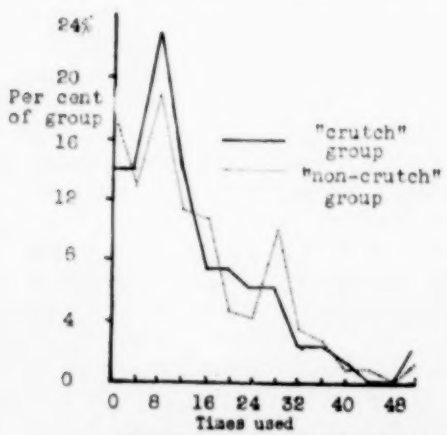


Figure 1. Per cents of the two groups using faulty procedures indicated number of times.

the solid line for the "crutch" pupils and the dotted line for the "non-crutch." It will be observed that the curves are markedly similar in shape and that they parallel each other rather closely. As measured by this method, then, there was little difference between the two groups in the amount of their faults.

But the limiting phrase "as measured by this method" proved to be a highly important one. The distribution table and Figure 1 did show general agreement in the use of faulty procedures, but being based upon gross measures, they could, it was clear, conceal real differences. For example, the distribution table revealed nothing regarding the relative number of different kinds of faults to which the two groups were subject. Evidence on this point was essential to a complete evaluation of the effect of the "crutch" upon learning. The first count seemed to indicate that the "crutch" children (perhaps because of the "crutch") employed a smaller number of different faulty procedures, their total being 34 as compared with 49 for the "non-crutch" children. Corrections (described in Part II) to remove the influence of size of group and of number of schools represented materially reduced the number of the "non-crutch" group to 36 or less. Thus the preliminary impression that the "crutch" had possibly performed a useful service in preventing wide variation in individual practice had to be modified, and again evidence was found to be lacking of any large amount of dissimilarity in the degree to which "crutch" and "non-crutch" children made use of faulty procedures.

Further study, however, produced other measures which, when applied, led again to changes in the evaluation of the worth of the "crutch." Neither of the two kinds of comparisons which thus far had been made had resulted in a single precise measure of the relative use of faulty procedures. For this purpose the obvious measure seemed to be an average. It was found that the "crutch" children on the average (median) had used faulty procedures 11.5 times and the "non-crutch" children, 12.5 times. When the types of faults on which the comparison was based were reduced to the 28 kinds which were found in common in the work of both groups, the averages became 11.5 and 10.5, respectively, for

"crutch" and "non-crutch" children. It now began to look as if the "crutch" instead of shielding children from faults might actually have exposed them excessively to faults.

Another kind of average was tried out. Whereas the first average was a measure of the number of times faulty procedures were employed regardless of the number of children who used them, this second average was specifically designed to take account of the number of users in each group. A distribution was first made for each group in such a way as to show the number of pupils who had employed each of the various faulty procedures. For these two distributions averages (medians) were computed to give the average number of pupils in each group who had used the different faulty procedures. These averages were then changed to per cents of each group in order to correct for differences in size. The resulting measure of 3.6% for the "crutch" group is interpreted thus: on the average the different faults of the "crutch" children appeared on the papers of 3.6% of the members of that group. The corresponding measure for the "non-crutch" group was 2.1%. The advantage of 1.5% for the "non-crutch" children remained when only the 28 common types of faults were included in the comparison (median for "crutch" children 4.8%, for "non-crutch" 3.3%).[1]

The nature of the final conclusion to be drawn from the comparison of the two groups on their use of faulty procedures thus depended upon which of the measures developed seemed to be most valid. On the basis of the distribution table and of an analysis of the variety of faults characteristic of the two groups little difference could be found, and the conclusion would be that the "crutch" apparently did not tend to involve its users in faulty procedures. On the basis of the averages (number of times faults appeared and number of children subject to faults) the "non-crutch" children seemed to be very slightly superior (though not reliably so), and the conclusion would be that the "crutch" had been a source of possible harm.

Use of particular types of fault. Thus far, conclusions were available only from comparisons which neglected the nature of faulty procedures. Even if there had not from the first been the intention to do so, the conflicting conclusions from the general comparisons would have made it imperative to study possible differences in the faults which were characteristic of the two groups of children. The next step seemed to be to compare the "crutch" and the "non-crutch" pupils as to the manner in which they had employed particular kinds of faulty procedure.

An examination of the data regarding the use of the 55 faults found on the papers of the 245 children at once revealed the fact that many could be eliminated from further consideration because of their infrequent occurrence. Two very lenient criteria, the application of which could not possibly lose anything of importance, were adopted as a means of selecting the particular faults which would be studied. In order to be included a given fault had to satisfy at least one of two sets of conditions: first, an average frequency of use of .20 times in one or both groups of children; second, a use of 7.5% by the children in one or both groups. A total of 32 faults failed to meet these requirements, and the comparison of the groups was then restricted to the 23 types of faults which remained.

The first test of different usage in the groups was that of relative number who had employed the separate faults (Part II, Table VI). For only five faults, here simply designated as faults D, G, M, U, and V, did there seem to be even a possibility of difference.

The second test was that of the average number of times the particular faulty procedures were used (Part II, Table VII). On this basis none of the previously rejected eighteen could be re-claimed for study as representing differences between the groups, nor could any one of the five which had met the first test meet the second as well.

A third test was then applied, a test which carried the study into the separate schools as a means of determining whether the five faults which had gotten by the first hurdle were actually characteristic of the group which was found to employ it more commonly. Faults D, M, and V were more popular among the "non-crutch" children. Faulty procedure D was found in use in only three

1. As a matter of fact the differences of 1.5% were not statistically significant, as shown in Part II.

of the five "non-crutch" schools. Its use was therefore not typical of "non-crutch" children, and it could not be accepted as a real difference between "non-crutch" and "crutch" children. Application of the same test showed that none of the other four faults (G, M, U, and V) was genuinely peculiar to the group in which it was found more commonly.

There remained a fourth test, the most crucial one, which would have been applied if this had been necessary, namely, a test of the functional relationship of the fault to use or non-use of the "crutch." The first three tests were purely statistical; they could do no more than show whether a given fault appeared more or less frequently among the "crutch" or the "non-crutch" children. Conceivably a faulty procedure might have been accidentally associated with the use of the "crutch" without its being caused by the "crutch" or even closely related to it. In other words, if a given faulty procedure had met the first three tests it would still have had to pass this one of functional connection. But no faulty procedure managed to get by the first three tests.

Conclusion to (b). The weight of the evidence, as obtained from the preponderant number of comparisons attempted, was that the "crutch" as it was employed in schools 1 and 2 represented a set of habits which were self-contained and relatively autonomous, a system which allied itself with or remained independent of other sets of habits, such as those involved in faulty procedures, as the circumstances of learning dictated. Certainly there was in the various sorts of analyses attempted little to support the contention that use of the "crutch" confused children and indirectly impeded their development of the ability to deal with proper fractions.

(c) Relationship of use of the "crutch" to accuracy in adding proper fractions

A third objection. Still, the opponent of "crutches" might not be convinced that children could without serious consequences be taught to add proper fractions through the agency of the elaborate form employed for that purpose in schools 1 and 2. This device might not become fixed as a permanent habit, and it might not interfere with learning in the manner in which its effect in this regard was here studied. But, he might contend, the "crutch" might exert its dire influence in some other way which would reveal itself in lowered ability to add fractions.

Here was another challenge, a vital test, which had to be met by the "crutch." Did the children who had been taught to employ the form fail to develop satisfactorily the ability to add proper fractions? Again data were at hand for an evaluation of the objection, this time in the scores on the fraction test.

Comparison of test scores. Test scores made by "crutch" and by "non-crutch" children were arranged in frequency distributions, and the median scores were calculated for the two groups (Part II, Table VIII). The average score for the "crutch" children was found to be 6.9 and that for the "non-crutch" children, 7.3. The difference, though too small to be taken seriously, was in favor of the "non-crutch" group, and on this account seemed to support the position of the opponents of "crutches."

A check was made on this conclusion by studying the comparative scores of the groups in a different way. Credit was allowed for answers to examples which were correct except for failure to reduce answers to lowest terms, and the new test scores were arranged in distributions (Part II, Table IX). The median for the "crutch" children rose to 12.8 while that for the "non-crutch" group rose to 14.2. Here was an even larger difference in favor of the "non-crutch" children, and furthermore a difference too large to be disregarded as negligible in amount. Something apparently had happened in the change from the one method of scoring to the other which had proved disastrous to the "crutch" children.

An attempt to account for differences. A search was instituted to isolate this disturbing factor, not merely in the hope of accounting for the loss suffered by the "crutch" children, but more in the hope that through the search something might be discovered concerning the relationship between use and non-use of the "crutch" and efficiency in adding fractions.

A comparison was first made of the relative success of the two groups of children

on the three types of example in the fraction test. No difference was observable in the case of examples 1-5 (common denominator in both fractions) or in the case of examples 10-15 (common denominator not given); but important differences could be detected in the case of examples 6-9 (common denominator in one fraction). Here the "crutch" children were distinctly inferior.

Just why the "crutch" children should have been so markedly less successful with these four examples was not at all clear, but it was even less clear that their relative loss in efficiency was to be traced to the influence of the "crutch." An explanation was first sought in the comparative manner in which these children employed faulty procedures in the three types of example. Two facts were uncovered: first, these children employed a greater variety of faulty procedures in these examples and, second, they employed them more frequently. They were especially prone to add the given denominators in order to secure the common denominator (instead of taking the common denominator as it was given in one fraction) and to make numerous errors in treating the numerators (being especially apt to add the numerators as given without changing them in terms of the common denominator). These facts were undeniable, but at the same time they failed entirely to require as their explanation some property of the "crutch" which necessarily brought about the situation as it was found to exist in the two schools which provided instruction in that device.

Conclusion to (c). The third of the three objections to "crutches" seemed, in the case of the particular "crutch" under observation, to be the soundest. The "non-crutch" children were, on the average, superior to the "crutch" children in adding proper fractions, although the amount of this superiority was too slight to meet the usual statistical requirements for reliability. Differences of an instructional sort other than those directly associated with the teaching of the "crutch" might explain the superiority of the "non-crutch" children fully as well as the fact that they did not employ the device in adding fractions. These other explanations for the difference, which were not here canvassed, seemed the more reasonable since no functional relation could be discovered between the nature of the "crutch" on the one hand and the special aspects of the ability to add fractions in which the "crutch" children were the weakest.

SUMMARY OF PART I

Taken as a whole the case against the "crutch" is not very strong. One by one, three objections to "crutches" in general have been raised against the particular one in use in schools 1 and 2, and none of these objections was found to be valid. Failure to discover reasons not to teach the "crutch" may readily be interpreted to signify that the "crutch" should be taught. Such an inference is, however, unwarranted. On the contrary, the whole study is partial and, in one sense of the word, negative. A great many aspects of the problem have not even been touched upon, and the ones which have received attention are those which had their origin in possible weaknesses of the device. Before a recommendation could be given this "crutch", it would be necessary to establish three facts: first, that it does not subject children to harmful influences other than those here considered; second, that it possesses distinct advantages which facilitate learning; and third, that it is at least as effective as other available methods of accomplishing the same end. The purpose of the study was not, however, to prove the worth of the "crutch", but rather to evaluate the device in so far as was feasible in terms of the available data. In other words the study is illustrative rather than final; it is an example of an attempt to substitute quantitative research for opinion in a field in which opinion still is too commonly the determining factor.[1]

1. The implications of the study for the psychology of learning and for the teaching of arithmetic are set forth in an article entitled "The Place of 'Crutches' in Instruction" soon to appear in the Elementary School Journal.

Part II

STATISTICAL EVIDENCE

Purpose and nature of Part II. The introductory material preceding Part I informed the reader of the purpose of the study as that of evaluating an arithemtic "crutch" from three points of view, reported the subjects to be 245 fifth grade children enrolled in seven schools, described the test on which data were secured, and illustrated the form of the "crutch" in connection with one example. Part I traced the course of the investigation, setting forth in detail the steps taken and giving the reasons therefor. In Part II very little of the account of the study will be repeated, only as much as may be necessary in presenting some of the basic data and in explaining such measures and interpretations as need clarification.

Certain definite limitations attach to the study as a whole. Chief among these is that fact that, due to the nature of the data collected and to the conditions under which they were obtained, the method of evaluation is largely statistical and analytical. For the adequate evaluation of the "crutch" a control-group experiment would have been much superior since it would have permitted immediate management of the factors which contributed the data. Under these circumstances it would have been possible to keep the "crutch" under direct observation and to note the character of its relation to learning in the process of the learning. The method of the present study is, by contrast, an ex post facto approach. That is to say, it represents a survey made after the factors had operated to produce certain results which then had to be scrutinized to detect possible relationships between factors and data. Yet, while these limitations must be frankly admitted, they do not by any means invalidate the findings. On the contrary, the conclusions are advanced with considerable confidence in their trustworthiness.

(a) Persistence of the "crutch." The first of the three separate studies, that dealing with the persistence of the "crutch", was confined to the pupils of schools 1 and 2. On the application of the fraction test to the 35 children in school 1 during the last week of January, 1930, it was found that these pupils used the "crutch" varying numbers of times. Nine did not use it at all; thirteen used it 10 times; and the remaining thirteen used it 15 times, in all of the examples. The manner in which use of the "crutch" varied within the group of 35 children suggested the hypothesis that they had tended to neglect and eventually to discard the "crutch" as it ceased to serve any purpose. In an attempt to check this hypothesis, as well as to secure further light on the nature of the persistence of the form, a number of measures were collected of both general and arithmetic ability. These measures for the 31 of the 35 children who took all the tests have been assembled in Table I. This table is read as follows: Subject 1 belongs to the group of children who did not use the "crutch" on the fraction test. His age was 142 months, his mental age 129 months, and his I.Q. 83. During the preceding year of two terms he had averaged 57.5 in arithmetic, according to his teacher's marks. He secured a score of 14 on the arithmetic problem-solving test in the National Intelligence Test (which was given to secure measures of M.A. and I.Q.). On the fraction test he solved five of the fifteen examples correctly and made the same score when credit was allowed for unreduced answers. The means for the three sub-groups of children are given in the rows following the three parts of the table. Thus, the means for the 0-group were 142.0 months in C.A., 140.5 months in M.A., and so on.

The separate items in Table I (for example, the C.A. of 139 for Pupil 31) possess little meaning; they are therefore in a form which is hardly conducive to understanding and interpretation. These data have, consequently, been re-worked with the results which appear in Table II. The procedure followed in deriving this table requires explanation. First, the 31 children, regardless of the number of times they employed the "crutch", were ranked with respect to the one trait of C.A. and were then divided into five groups of approximately equal number. Thus, there were six children (pupils 2, 6, 42, 61, 65, 66) whose ages fell within

TABLE I

DATA CONCERNING PUPILS OF SCHOOL 1, WHO USED FRACTION "CRUTCH" VARYING NUMBERS OF TIMES IN TEST 1

Pupil	Frequency of Use of the "Crutch"	General Measures			Measures of Arithmetic Ability			
		C.A. in Mos.	M.A. in Mos.	I.Q. in Mos.	Average of Teachers' Marks	N.I.T. Problem-Solving Score	Fraction Test: Number Exactly Right	Fraction Test: Number Right Except Reduction
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	0	142	129	83	57.5	14	5	5
2	0	158	152	98	59	14	1	1
3	0	140	138	99	57*	14	3	5
4	0	132	151	114	89.5	20	6	6
5	0	128	138	110	78*	16	9	15
6	0	160	152	95	86*	12	9	15
7	0	144	131	91	69	12	6	15
8	0	132	133	101	81	14	10	12
Mean		142.0	140.5	98.8	72.2	14.5	6.1	9.3
31	10	139	160	115	85.5	16	10	10
32	10	130	181	139	94	18	10	11
33	10	148	149	101	97	22	10	10
34	10	143	157	109	87.5	20	15	15
35	10	120	141	117	80.5	18	9	13
36	10	132	130	98	84	10	12	15
37	10	144	137	95	54	14	6	15
38	10	133	122	91	79	16	11	11
39	10	144	146	101	79	14	9	11
40	10	132	138	104	70	12	10	14
41	10	120	147	132	77*	14	12	12
42	11	152	137	90	62	6	9	10
Mean		136.4	145.3	107.7	79.1	15.0	10.3	12.3
61	15	161	138	85	84	16	14	15
62	15	132	132	100	69	16	7	15
63	15	120	159	132	89	18	12	13
64	15	136	136	100	76	20	4	4
65	15	149	150	101	64	16	8	10
66	15	170	129	76	60	10	5	11
67	15	144	114	79	70*	12	7	15
68	15	132	129	98	82	12	7	15
69	15	120	132	110	88	8	13	14
70	15	145	146	101	60	12	6	15
71	15	136	154	113	93*	16	12	15
Mean		140.5	138.1	99.5	75.9	14.0	9.1	13.1

*Teachers' marks thus indicated represent one term's work; all others are based on the average of two terms' work in arithmetic.

the limits of 149-70 months; these were the oldest children and are here included in the first or highest fifth. Six other children (pupils 7, 33, 37, 39, 67, 70), whose ages lay between 144-48 months, constitute the second fifth, and so on. The data for M.A., I.Q., average mark in arithmetic, etc., were all separately transmuted in a similar manner, and the 31 children were divided into approximately equal fifths on each scale.

1. This division into fifths was, as stated, only approximate. The theoretic number which should have appeared in each fifth was 6.1 (31 divided by 5). Actually in only eight of the 35 fifths in Table II are there less than 6 individuals or more than 7, seven of these being 5. One important variation from the practice of dividing into fifths appears in the last set of three columns. Twelve of the 31 children made perfect scores on the fraction test when credit was

TABLE II

SCHOOL 1, GRADE V. COMPARISONS OF THREE GROUPS OF USERS OF THE "CRUTCH" WITH RESPECT TO CERTAIN MEASURES OF DEVELOPMENT AND ARITHMETIC ABILITY

Fifths of Pupils with Respect to Trait Named at Head of Each Set of Three Columns	C.A.	M.A.	I.Q.	Average Mark in Arith.	Score in Arith., N.I.T.	Score-Fraction Test	
						Answers Entirely Right	Answers Right Except Reduc.
	15 10 0	15 10 0	15 10 0	15 10 0	15 10 0	15 10 0	15 10 0
First (highest)	3 1 2	2 3 2	1 4 1	3 2 1	2 4 1	5 3 0	6 3 3
Second	2 3 1	2 3 1	2 2 1	2 3 1	4 2 1	0 5 1	
Third (middle)	2 2 2	1 2 2	4 2 1	1 4 2	0 3 4	0 3 2	2 3 1
Fourth	2 3 2	3 2 2	1 2 3	3 2 1	2 1 2	4 1 2	2 6 0
Fifth (lowest)	2 3 1	3 2 1	3 2 2	2 1 3	3 2 0	2 0 3	1 0 4
Total	11 12 8	11 12 8	11 12 8	11 12 8	11 12 8	11 12 8	11 12 8

Table II consequently shows the distribution of the 31 children, first, according to their position in the various fifths for the different traits (C.A., M.A., etc.), and second, according to the number of times they used the "crutch" in the fraction test (15, 10, and 0). Thus, one notes from the table that of the oldest (highest) fifth, three pupils used the "crutch" 15 times; one used it 10 times, and two used it not at all. Of the seven most mature mentally (M.A.), two used the "crutch" 15 times, three 10 times, and two 0 times.

With the data in this form it was now possible to study correspondence between any of the separate measures and the number of times the "crutch" was used on the test. For example, any marked relation among these 31 children between age-level and continued use of the "crutch" would show itself in a tendency for the data to be concentrated in certain sections of the part of the table which is devoted to C.A. Thus, if the older children were the ones first to abandon the use of the "crutch" that condition would be revealed by the presence of a disproportionately large number of individuals in the upper right and the lower left corners of the section and by a disproportionately small number of individuals in the upper left and lower right corners. On the other hand, if the reverse of this tendency obtained and the older children tended to rely increasingly on the "crutch", this condition would be evidenced by a clustering of cases in the upper left and the lower right corners and by a lack of cases in the upper right and the lower left corners.

There is no marked concentration in any part of the C.A. section of Table II. On the contrary, the cases are fairly evenly distributed, indicating the absence of any type of relationship between C.A. and persistence in the use of the "crutch." If the table is read vertically by columns, instead of horizontally by rows, the same conclusion is seen to be the only tenable one. Thus, of the eleven children who used the "crutch" 15 times three were among the oldest six in the grade, two were in the second fifth, and two also were in each of the three lowest, or youngest, fifths. The cases in the columns for the 10- and the 0-groups are similarly distributed.

The method of reading the data for M.A. and I.Q. in Table II is the same as that for C.A. Relationship between persistence in the use of the "crutch" and these items of general ability would be similarly manifested by the distribution of cases within the

(Footnote continued)
allowed for unreduced answers, and it was therefore necessary to call them the two highest fifths, no difference between them being recognized.

appropriate parts of the sections. With respect to both M.A. and I.Q., as was the case with respect to C.A., the 31 children are rather uniformly scattered and the conclusion with regard to them is like that for C.A. in being negative. In this group of children there was no marked association between measures of these traits and continued use of the form; in other words, the "crutch" persisted or disappeared from use quite apart from the single fact of differences in mental maturity or in brightness.

The four measures of arithmetic ability in Table II represent three distinct aspects of the total ability. The first (average mark in arithmetic) is the most general. It represents the teacher's evaluation of the year's standing [1] of each child in all those arithmetical processes which were taught in grade IV, and hence does not include ability to add proper fractions. The second measure is based on the problem-solving test in the National Intelligence Test. These scores, therefore, refer to a particular arithmetic ability rather than to the composite comprehended in teachers' marks and refer, furthermore, to an ability which is supposedly quite different from that involved in adding fractions. The last two measures were obtained from the fraction test and are therefore measures of the specific ability under investigation.

The data for the first two measures of arithmetic ability may be dismissed with the brief statement that they reveal no explanation at all for continued use of the "crutch." The last two measures deserve more detailed consideration because of their being measures of the ability to add fractions. When credit was allowed only for completely correct answers, five of the fourteen who made the highest scores on the fraction test used the "crutch" 15 times and thirteen of them 10 or more times. On the other hand, six of the twelve who made the lowest scores used the form 15 times, and seven, 10 or more times. Excessive dependence on the device was, then, almost as apt to be associated with inability as with marked ability to add proper fractions. When credit was allowed for answers correct except for failure to reduce, a similar association was revealed. It is true that half of those in the highest two fifths on the fraction test used the "crutch" 15 times and that only one-fourth (three out of thirteen) in the lowest two fifths in the test used the form this number of times, but this apparent difference is offset by a corresponding reversal for those who used the form 10 times and by a fairly even distribution of cases who did not use the "crutch" at all.

The study thus far had been concerned with the relationship between retention of the "crutch" on the one hand and measures of C.A., M.A., I.Q., and of various aspects of arithmetic ability, each taken separately, on the other hand. The conclusion, that no one of these traits alone was closely associated with use or disuse of the form, might be true, at the same time that there might have been a relationship between persistence of the "crutch" and certain combinations of items. This hypothesis was tested in the case of two of the more promising combinations, namely, M.A. and general arithmetic ability, and I.Q. and general arithmetic ability. The method of study is illustrated in Table III, which is interpreted as follows: the seven children who were most mature mentally (highest fifth, M.A. 152-81 months) are distributed according to their standing in general arithmetic ability and according to their use of the "crutch." Thus, there were three of the seven who were also in the highest fifth of arithmetic ability. Two of the three used the form 15 times, and the other one, 10 times. There were also three children who ranked in the highest fifth in M.A., in the second highest fifth in arithmetic ability. Of these, two used the form 10 times, and the other 0 times. The last of the seven most mature mentally was in the poorest fifth in arithmetic, and he made no use of the form. The data for the other fifths in M.A. are read in similar fashion. An analysis of the data in the sections for the separate fifths of the table is space-consuming and may easily be made by the reader himself. In some cases (for example, in the data for the highest fifth in which excessive use of the "crutch" seems to be related both to high M.A. and to high arithmetic ability) relations might be found, but these relations would not hold in

1. Except in the case of six children (see Table I) for whom the records showed marks for only one of the two items.

TABLE III

SCHOOL 1, GRADE V. RELATION OF PERSISTENCE OF THE "CRUTCH" TO VARIOUS COMBINATIONS OF M.A. AND GENERAL ARITHMETIC ABILITY

Fifth in M.A.	Fifth in Arith. (Teachers' Marks)	Frequency of Use of the "Crutch"			Total
		15	10	0	
First (highest) 152-81 months	First 88 - 97	2	1	0	
	Second 82 - 87	0	2	1	
	Third 76 - 81	0	0	0	
	Fourth 62 - 70	0	0	0	
	Fifth 54 - 60	0	0	1	7
Second 146-51 months	First	0	1	1	
	Second	0	0	0	
	Third	0	2	0	
	Fourth	1	0	0	
	Fifth	1	0	0	6
Third (middle) 138-41 months	First	0	0	0	
	Second	1	0	0	
	Third	0	1	1	
	Fourth	0	1	0	
	Fifth	0	0	1	5
Fourth 131-37 months	First	1	0	0	
	Second	0	0	0	
	Third	1	0	1	
	Fourth	1	1	1	
	Fifth	0	1	0	7
Fifth (lowest) 114-30 months	First	0	0	0	
	Second	1	1	0	
	Third	0	1	0	
	Fourth	1	0	0	
	Fifth	1	0	1	6
Total		11	12	8	31

other fifths of the table (for example, excessive use of the "crutch" is also characteristic of those who were low in M.A. and in arithmetic ability). The only conclusion to be had from a consideration of all 31 cases is that whatever may have been the reason for their retaining or relinquishing use of the "crutch", that reason is not to be found in any peculiar combination of M.A. and general arithmetic ability.

A table similar to Table III was constructed to permit the study of the relation between I.Q. and general arithmetic ability, combined, and retention of the "crutch." This table, here omitted, likewise yielded negative results; persistence of the "crutch" was not to be accounted for in terms of some special combination of this pair of items.

The fraction test was given to the 45 children in grade V of school 2 during the first week of February, 1930. Table IV contains the data which were collected for these pupils. The table is divided into two parts, the upper half containing the measures for the children who did not employ the "crutch" and the lower half containing the measures for the children who used it 15 times. Unlike Table I for the children in school 1, Table IV does not include the measures for the problem-solving test in the National Intelligence Test.

No exhaustive analysis like that for school 1 was undertaken for school 2. Three reasons were responsible for this omission. In the first place, the school organization into upper and lower sections of the grade

TABLE IV

SCHOOL 2, GRADE V. DATA CONCERNING PUPILS WHO USED FRACTION "CRUTCH" VARYING NUMBERS OF TIMES IN TEST 1

Pupil	Use of "Crutch" Test 1	General Measures			Measures of Arithmetic Ability		
		C.A.	M.A.*	I.Q.*	Teachers' Marks	Test 1--Number Exactly Right	Test 1--Number Right, Except Reduction
Section B							
101	0				67	5	11
102	0	130	160	123	80	2	12
103	0	127	139	109	64	5	5
104	0	133	134	101	55	0	0
105	0	124	140	112	77	6	14
106	0	130	136	104	65	2	8
107	0	139	148	106	59	6	15
108	0	126	165	130	89	8	14
109	0	124	186**	152**	93	7	14
110	0	124	131	105	68	7	15
111	0	137	163	118	100	14	15
112	0	125	131	104	60	2	12
113	0	124	142	114	76	7	15
114	0	116	142	122	70	7	14
115	0	121	158	130	76	0	0
116	0	128	136	106	35	1	15
117	0	127	162	127	88	4	14
118	0	127	142	111	43	1	12
119	0	135	155	114	63	7	15
120	0	126	122	96	69	8	15
121	0	125	151	120	95	7	15
122	0	127	142	111	48	7	9
123	0	118	163	138	94	15	15
Mean		127.0	147.6	116.1	71.0	5.5	11.8
Section C							
131#	0	128	127	99	96	6	12
132	15	126	151	119	49	0	0
133	15	102	132	129	60	0	8
134	15	134	132	98	78	2	12
135	15	125	134	107	99	8	15
136	15	134	146	108	66	4	11
137	15	120	118	98	29	0	0
138	15	126	130	103	67	1	12
139	15	134	139	103	50	1	13
140	15	127	162	129	97	3	14
141	15	122	142	116	52	7	15
142	0	122	138	113	60	3	14
143	15	125	133	106	61	3	9
144	15	132	154	116	54	0	0
145	15	143	119	83	55	5	9
146	15	118	186**	158**	96	6	12
147#	0	140	135	96	68	1	1
148	15	121	131	108	53	5	15
149	15	122	113	92	49	1	2
150	15	148	116	78	46	2	10
151	15	121	134	111	57	4	8
152	15	130	131	101	90	5	12
Mean		126.6	134.5	108.8	63.4	3.0	9.6

*M.A.'s and I.Q.'s were obtained through the use of the National Intelligence Test, which was administered by the writer.

**These measures are estimated. The highest score for which an M.A. is given is the score of 135. These two children scored higher than this.

#These Children failed of promotion the preceding quarter, and were required to repeat the work of the C-section. There scores are, on this account, omitted in calculating the measures for the C-group as a whole.

(B and C, respectively) which were separated by three months of schooling, paralleled and seemed to account for the difference in use of the "crutch." In the second place, there were no differences within either section with regard to the use of the form, a condition which made such a study unnecessary, as it was necessary in the case of school 1. In the third place, the number of children in the two sections would have been too small to have justified the types of distinction made in the case of school 1 even if such an analysis had been deemed advisable. It is, however, interesting to compare the means of the 0-children (B-section) with the means of the 15-children (C-section). The former group, though of the same average C.A., were more than a year older mentally and were considerably brighter (an average margin of more than seven points in I.Q.). They thus seem to have been selected for Section B largely on the basis of intellectual factors. In arithmetic ability in general and in scores on the fraction test they showed the superiority that would have been expected from their greater mental ability and from their larger amount of schooling.

The rest of the search for the factors underlying the persistence of the "crutch" is best detailed in the narrative of Part I. Statistical analysis played an exceedingly small role in that search. Solution of the problem came through simple comparisons, through the results of re-testing, and finally through conversation and correspondence with the classroom teachers. On this account, there is no further reference in Part II to the first problem investigated. The following section of Part II is devoted to the second problem, namely, the degree to which use of the "crutch" interfered with learning to add fractions by involving its users in faulty procedures.

(b) Use of faulty procedures. As explained in section (b) of Part I, an attempt was next made to determine whether the "crutch" had interfered with the development of ability to add proper fractions. The means adopted to measure this interference with learning was to note whether use of the "crutch" was related in any close way to use of faulty procedures (errors, roundabout processes, etc.) in adding fractions. A list of 55 such faults had been collected from the 245 subjects in the investigation through a personal conference with each child.

In order to determine whether or not use of the "crutch" had led to an excessive dependence on faulty procedures in general or on some particular faulty procedures the 80 children in the two "crutch" schools were compared with the 165 children in the five "non-crutch" schools. In other words the "non-crutch" schools served as a check group. They could serve this purpose only on the condition that the two groups of schools (and of children) were alike in all essentials except that of instruction in the "crutch." There is considerable reason to believe that the requirement of similarity was met. In the first place, both sets of schools used the same arithmetic texts, followed the same course of study, and allotted the same amount of time to the subject. In the second place, economically and socially the communities served by the two sets of schools were about equally representative of the total population served by the seven schools. One of the "crutch" schools was located in a suburban area slightly below average in economic circumstances and social standing. The other of the "crutch" schools was situated in a slightly better-than-average city district. The five "non-crutch" schools were distributed as follows: one in a very poor suburban area, one in an average suburban area, one in a superior suburban area, one in an average city area, and one in a slightly superior city area. These facts, which show likeness both in the general features of arithmetic instruction and in pupil material, would seem to justify the comparisons which were made between "crutch" and "non-crutch" children.

The first comparison had to do with the number of times faulty procedures were used. For this purpose Table V was prepared. It is read in the following manner (fourth row): two children in schools 3-7 ("non-crutch" schools) exhibited 40 to 43 faults in their work as compared with one such child in schools 1 and 2 ("crutch" schools). These numbers of children represented respectively 1.2 per cent and 1.3 per cent of their groups. A total of 3.6 per cent of the "non-crutch" and of 3.8 per cent of the "crutch"

TABLE V

COMPARISON OF SCHOOLS 3-7 ("NON-CRUTCH" SCHOOLS) WITH SCHOOLS 1 AND 2 ("CRUTCH" SCHOOLS) WITH RESPECT TO FREQUENCY OF USE OF ALL TYPES OF FAULTY PROCEDURES COMBINED

Frequency of Faulty Procedures	Number of Pupils		Per cent of Whole Group		Cumulative Per Cents	
	Schools 3-7	Schools 1 & 2	Schools 3-7	Schools 1 & 2	Schools 3-7	Schools 1 & 2
More than 48	2*	2**	1.2	2.5	1.2	2.5
48	0	0	0.0	0.0	1.2	2.5
44	2	0	1.2	0.0	2.4	2.5
40	2	1	1.2	1.3	3.6	3.8
36	5	2	3.0	2.5	6.6	6.3
32	6	2	3.6	2.5	10.2	8.8
28	16	5	9.7	6.3	19.9	15.1
24	7	5	4.2	6.3	24.1	21.4
20	8	6	4.8	7.5	28.9	28.9
16	18	6	10.9	7.5	39.8	36.4
12	19	11	11.5	13.8	51.3	50.2
8	31	18	18.8	22.5	70.1	72.7
4	21	11	12.8	13.8	82.9	86.5
0	28	11	17.0	13.8	99.9	100.3

*Actual frequency of faulty procedures, 60 and 65.
**Actual frequency of faulty procedures, 54 and 67.

children used faulty procedures the same or a larger number of times. The data in the second pair of columns were utilized in Part I to construct Figure 1 in order to show graphically the relative distribution of pupils in the two groups with respect to the scale of frequencies given in the first column of Table V. The last two columns reveal the same facts through cumulative per cents. It will be noted that 28.9 per cent of both groups used faulty procedures 20 or more times and that approximately 50 per cent used faulty procedures twelve or more times. On this basis the two groups of children are seen to be more alike than different in the degree to which faults appeared in their work.

Even though it did permit a study of the two groups point for point on the scale, this type of comparison was not satisfactory. It entirely neglected the question regarding the variety of faults to which the groups were subject. Accordingly a count was made of the number of different faults which appeared on the papers of the "non-crutch" children and of the "crutch" children. The totals were found to be 49 and 34 respectively. The difference of 15 seemed to be evidence of considerable unlikeness, with the "crutch" children holding distinct advantage. Still the difference in number of faults might have been due to the larger size of the "non-crutch" group or to their enrollment in five instead of two schools. Checks were accordingly made of the influence of these two factors.

In order to secure a group of 80 "non-crutch" children who would be comparable with the 80 "crutch" children a sample of 80 was drawn from the 165 "non-crutch" children in such a way that the proportionate number contributed by each school was the same as its contribution to the total of 165. The number of different faults in the work of this sample of 80 "non-crutch" children was 38. Four other similar samples were selected, and their variety of faults was studied. The range of different types was from 36 to 40, giving a median of 37 for the five samples. This figure, 37, was a better measure than was the 49 originally obtained, of the number of different faults to which the "non-crutch" children were subject. It differs by only 3 from the corresponding figure for the "crutch" group.

Yet small as was this difference, it still failed to take account of the possible influence from the unlike numbers of schools represented in the two groups. In one of the two "crutch" schools 30 different faults were discovered; in the other, 32. In the "non-crutch" schools the following numbers were found: 19, 20, 36, 21, and 28. Actually, therefore, three of the five "non-crutch" schools used fewer different kinds of faulty procedure than did one of the "crutch" schools, and four of the five used fewer than did the other one of this pair. The larger total of 49, and the smaller one of 37 as given above, are thus seen to be a thoroughly unsatisfactory measures of the relative number of different faults to be found in the fraction habits of these "non-crutch" children. A valid measure for this group would tend to wipe out entirely their apparent difference from the "crutch" children and would serve to emphasize likeness rather than dissimilarity in respect to the variety of their faults.

The two preceding comparisons were between the groups as wholes. In order to express relative use of faulty procedures more precisely, the study was continued in terms of averages. To secure the first average, the individual children in the two groups were separately arranged in the order of the number of faults on their papers. Then for each array the median was determined. This proved to be 12.5 in the case of the "non-crutch" children and was interpreted to mean that the average "non-crutch" child used faulty procedures 12.5 times. The corresponding median for the "crutch" children was 11.5. It was already known that the two groups used different numbers of particular faulty procedures, and there was the chance that the above medians, which included all types of fault, might discriminate unfairly against one or the other of the groups. Accordingly those faults were eliminated which were peculiar to either set of children, and medians were computed on the basis of the 28 faults which were common to both groups. The new median for the "non-crutch" group was 10.5, and for the "crutch" group, 11.5, the advantage, though slight, having changed from the "crutch" to the "non-crutch" children.[1]

The second average used in the comparison of the groups was derived from the number of children who employed the various faulty procedures of their group rather than from the number of times these faults were exhibited. Tabulations were first made to discover the number of "non-crutch" children who had employed each of the faulty procedures which were used by their group. Then from this array the median was obtained, it being 3.5. The corresponding median for the "crutch" group was 2.9. In order to make the medians comparable, allowance had to be made for the number of children in the two groups. The medians were divided by 165 for the "non-crutch" group and by 80 for the "crutch" group. The resulting measures, 2.1 per cent and 3.6 per cent, represent the proportion of each group who on the average used the faulty procedures of their group. When the basis of comparison was again changed to the 28 types of fault which were common to the two groups, these figures became 3.3 per cent and 4.8 per cent.[2] On both bases the "crutch" children showed to a disadvantage.

To summarize, the two comparisons of a more general nature, that based on Table I showing number of times faulty procedures were used and that which took account of the variety of faults for the two groups, seemed to indicate that the "crutch" had had no harmful or retarding influence on learning. The two comparisons based on averages were opposed to such a conclusion, and gave some small amount of evidence in favor of the position that the "crutch" actually interfered with learning.

Up to this point the comparisons had neglected the nature of the different faulty procedures which were characteristic of the "non-crutch" and of the "crutch" children. The two groups could have been very much alike with regard to their use of faulty procedures in general and yet have been very unlike in the particular faults to which they were subject. It was not necessary to include in the comparisons all 55 types of

1. In the case of neither of the comparisons was the difference reliable. The ratio of the difference in medians to the $P.E._{diff.}$ was .41 when all faults were included, and was .79 when only the 28 common faults were considered.
2. The differences of 1.5 per cent were not statistically significant, the ratios of the difference of the per cents to $P.E._{diff.}$ being .91 and .39, respectively, in the comparisons described.

TABLE VI

COMPARISON OF SCHOOLS 3-7 ("NON-CRUTCH" SCHOOLS) AND SCHOOLS 1 AND 2 ("CRUTCH" SCHOOLS) WITH RESPECT TO THE NUMBER OF USERS OF PARTICULAR TYPES OF FAULTY PROCEDURES

Types of Faults	Number of Users		Percentage of Total		Difference in Percentage	P.E. of Difference*	Difference / P.E. Diff.
	Schools 3-7	Schools 1 & 2	Schools 3-7	Schools 1 & 2			
A	15	8	9.1	10.0	.9	8.5	.11
B	11	10	6.7	12.5	5.8	9.6	.60
C	5	2	3.0	2.5	.5	2.4	.21
D	15	2	9.1	2.5	6.6	3.9	1.69
E	84	50	50.9	62.5	11.6	22.4	.52
F	24	15	14.5	18.8	4.3	13.8	.31
G	1	2	.6	2.5	1.9	1.9	1.00
H	7	3	4.2	3.8	.4	3.4	.11
I	44	16	27.3	20.0	7.3	12.3	.59
J	3	2	1.8	2.5	.7	2.1	.32
K	23	8	13.9	10.0	3.9	9.8	.39
L	6	6	3.6	7.6	4.0	7.1	.56
M	16	..	9.7	....	9.7	3.5	2.77
N	13	4	7.9	5.0	2.9	4.9	.58
O	130	65	78.8	81.3	2.5	13.3	.18
P	2	1	1.2	1.3	.1	1.1	.09
Q	27	13	16.4	16.3	.1	12.6	.01
R	76	29	46.1	36.3	9.8	25.3	.38
S	19	7	11.5	8.8	2.7	8.1	.33
T	54	20	37.5	25.0	12.5	18.8	.66
U	...	1	....	1.3	1.3	1.1	1.21
V	20	..	12.1	....	12.1	4.3	2.80
W	5	3	3.0	3.8	.8	3.2	.25

*Formula for the Probable Error of the Difference:

$$.6745\sqrt{\left(\frac{p_1q_1}{N_1}\right)^2+\left(\frac{p_2q_2}{N_2}\right)^2}$$

in which "p" represents the per cent of one variable, "q" the per cent of the other, and "N" the number of cases. The figures "1" and "2" refer to the two series of scores.

fault. On the contrary, 32 could be disregarded from the outset because of the infrequency of their appearance in either group. The remaining 23 types were first studied in terms of the number of children who used them. The pertinent data are presented in Table VI, which is read: Fault A appeared on the papers of 15 of the "non-crutch" children and on the papers of 8 of the "crutch" children. These figures represent 9.1 per cent and 10.0 per cent of their respective groups. The difference in per cent is .9 point and the $P.E._{Diff.}$ is 8.5. The ratio of difference to $P.E._{Diff.}$ is .11. The last column of Table VI is the important one. In the case of only five of the faults (D, G, M, U, and V) is this ratio as large as 1.00.

Next the 23 faults were studied from the standpoint of the number of times they were used. The data for this comparison are found in Table VII, which is constructed like Table VI except for the basic unit: times used instead of number of users. In the case of none of these faults is the ratio of difference to $P.E._{Diff.}$ as much as 1.00.

Finally the five faults which of all 23 had seemed to show the largest amount of difference in number of users in the two groups were subjected to examination with a view to determining whether they were really typical of the group which used them more frequently.

TABLE VII

COMPARISON OF SCHOOLS 3-7 ("NON-CRUTCH" SCHOOLS) WITH SCHOOLS 1 AND 2 ("CRUTCH" SCHOOLS) WITH RESPECT TO MEAN FREQUENCY OF USE OF PARTICULAR FAULTS

Type of Fault	Total No. Uses		Mean No. Uses		Difference in Means	P.E. of Difference*	Difference / P.E.$_{Diff.}$
	Schools 3-7	Schools 1 & 2	Schools 3-7	Schools 1 & 2			
A	20	10	.12	.13	.01	.40	.03
B	106	71	.64	.89	.25	2.31	.11
C	41	1	.25	.01	.24	.99	.24
D	85	8	.52	.08	.42	1.23	.34
E	231	199	1.40	2.49	1.09	2.60	.41
F	69	25	.42	.31	.09	1.36	.07
G	1	16	.01	.20	.19	1.57	.12
H	100	41	.61	.52	.09	2.38	.04
I	51	27	.32	.34	.02	.59	.34
J	15	17	.09	.21	.12	2.91	.04
K	208	69	1.26	.86	.40	2.72	.15
L	8	34	.05	.43	.38	1.99	.20
M	16	...	.09	....	.09	.78	.10
N	15	4	.09	.05	.04	.54	.08
O	610	376	3.70	4.70	1.00	2.72	.36
P	5	27	.03	.34	.31	1.14	.27
Q	40	19	.24	.26	.02	.54	.36
R	377	132	2.29	1.65	.64	2.73	.23
S	42	13	.25	.16	.09	.88	.10
T	58	25	.35	.31	.04	1.68	.02
U	...	15	....	.19	.19	1.11	.17
V	275	...	1.67	....	1.67	3.03	.55
W	23	32	.14	.40	.26	1.62	.16

*Formula for the Probable Error of the Difference:

$$P.E._{Diff.} = \sqrt{P.E._{M_1}^2 + P.E._{M_2}^2}$$

Three faults were more popular among the "non-crutch" children. Fault D was, however, exhibited by only fifteen of the 165 "non-crutch" children, and these fifteen were confined to three of the five schools in the group. Fault M was found in the work of only sixteen "non-crutch" children, and fifteen of the sixteen were furnished by a single school. Faulty procedure V was employed by twenty "non-crutch" children who were drawn from the five schools as follows: ten, seven, two, one, none. None of these faults was therefore really characteristic of "non-crutch" children in general. Neither were Faults G and U characteristic of "crutch" children even though they did exhibit them more frequently than did the "non-crutch" children. Reference to Table VI shows that the first of these faulty procedures was used by only two children, and the second by only one child in the "crutch" schools.

A fourth test was still available had it been necessary, to detect difference between the groups in the types of fault they preferred. This test was one of the functional relation of the "crutch" to the fault. To have met the test a given fault would have had to be so closely related to the "crutch" that the use of the latter would have inevitably involved the user in the fault. The failure of any of the faults to pass the first three tests, or even two of the first three tests, made the fourth test superfluous.

All things considered, therefore, there seemed to be little basis for the belief that the use of the "crutch" had interfered with learning to add proper fractions either

by making its users excessively dependent upon faulty procedures in general or by leading them to adopt one or several of the different types of these faults.

(c) Relation between use of the "crutch" and accuracy in adding proper fractions. As a final check on the influence of the "crutch" on learning the "non-crutch" and the "crutch" children were compared as to their relative success on the fraction test. Scores on this test were taken as being at least indicative of the degree to which the two groups of children had developed ability in that arithmetical process in connection with which the "crutch" had been taught in schools 1 and 2. As measures of ability to add proper fractions the scores on the fraction test were not as satisfactory as they would have been, had it been possible at the time of the testing to have foreseen the use to which they were later put.

The fraction test had not been especially devised for the purpose for which scores on the test were now to be utilized. The test was short;[1] it was not timed; it yielded scores only in terms of accuracy, and it did not contain very many difficult examples. Then too the kinds of measures secured were distinctly limited. In order to furnish adequate data for a final answer to the question proposed, measures were required of other aspects of the ability to add proper fractions than those aspects which were covered by the test, such as skill in adding such fractions in verbal problems, the extent to which skill in adding proper fractions carried over to the addition of improper fractions and mixed numbers, etc.

In spite of these limitations, however, there seemed to be some value in studying the relative success of these two groups of children in taking the fraction test. In case that extreme differences were found between the groups, the results of the study would at least foreshadow, even if they did not decisively reveal, the general nature of the relationship of the "crutch" to proficiency in the particular arithmetical skill under investigation. Thus, if the "crutch" children were found to be vastly superior to the "non-crutch" children in adding the test examples, then it should be safe to conclude that the "crutch" had worked no great hardships upon the former, such hardships as the opponents of this "crutch" might easily conceive on theoretical grounds. And, on the other hand, if the "crutch" children proved to be greatly inferior to the "non-crutch" children, it would hardly be possible to claim for the "crutch" many of the values which its proponents might be inclined to advance for it.

Table VIII contains a distribution of the scores for the two groups of children when credit was allowed only for completely correct answers. Thirteen of the "non-crutch" children made perfect scores, as did four of the "crutch" children, and so on. The median score for the "non-crutch" group was 7.3; for the "crutch" group, 6.9. The difference of .40 in medians was found to be unreliable.

The method of scoring was then made more lenient, and credit was allowed for answers correct except for failure to reduce. Table IX presents the distribution of scores on this basis. It is read like Table VIII. The difference in medians was now 1.4 in favor of the "non-crutch" group, a difference which met reasonable requirements for statistical reliability.

An attempt was next made to discover through an analysis of the relative achievement of the two groups on the three types of example in the test a possible explanation for the comparative loss of efficiency on the part of the "crutch" children. The method adopted for this study may be explained from Table X. Of the first type of example (denominators already similar) there were five. In the "non-crutch" schools 119 children computed all five correctly, 25 secured correct answers for four, six for three, none for two, four for one, and eleven for none. The 165 children in these schools secured a total of 717 correct answers out of a possible 825 solutions (165 times 5), making a per cent of 86.9. The corresponding figure for the 80 "crutch" children on these examples was 85.8 per cent.

1. In spite of this fact the test was surprisingly reliable. Scores on the eight odd-numbered examples correlated with scores on the seven even-numbered examples .83. The reliability of the 15-example test was therefore .91, after the coefficient had been stepped up by the Spearman-Brown formula. Application of this formula seemed to be justified by the fact that the statistical requirements for its use were met as well as by the fact that the content of the test-halves was closely homogeneous.

TABLE VIII

COMPARISON OF SCHOOLS 3-7 ("NON-CRUTCH" SCHOOLS) AND SCHOOLS 1 AND 2 ("CRUTCH" SCHOOLS) WITH RESPECT TO SCORES ON THE FRACTION TEST; CREDIT ALLOWED ONLY FOR COMPLETELY CORRECT ANSWERS

Scores in Terms of Number of Completely Correct Answers	Schools 3-7 ("Non-crutch" Schools)	Schools 1 & 2 ("Crutch" Schools)
15	13	4
14	16	2
13	7	1
12	4	6
11	7	1
10	9	5
9	8	6
8	8	4
7	16	10
6	15	8
5	10	8
4	9	4
3	10	4
2	13	5
1	7	6
0	13	6
Total Cases	165	80
Median Score	7.3	6.9
Q. Same	4.0	3.0
$P.E._{Md.}$	.39	.42
Difference in Medians	.40	
$P.E._{Diff.}$	.57	
$\frac{\text{Difference}}{P.E._{Diff.}}$	.70	

The comparisons afforded by the last line of the table reveal the fact that the "non-crutch" children were very slightly superior to the "crutch" children in the first and third types of example (the third type requiring that the common denominator be found),[1] but that they were distinctly superior in the case of the second type (one of the denominators being the common denominator). Here the difference of 10.4 per cent was 1.4 times the $P.E._{Diff.}$. On this kind of example in the addition of proper fractions the "crutch" children were no match for the "non-crutch" children (that is, of course, within the limits of the reliability of the observed difference).

In the attempt to discover the reason for this special deficiency of the "crutch" children the form was studied in relation to the processes required to solve the type-two examples. Analysis was here resorted to in the hope of isolating those properties of the "crutch" which would make its users encounter unusual difficulty with examples of this kind. No such properties could be found. A second type of analysis was then undertaken in the effort to determine whether the "crutch" children tended in type-two examples to vary the number, or kinds, or frequency of their faulty procedures. The search revealed the fact that in this respect, that of faults, the work of the

1. Difference in type one, 1.1; $P.E._{Diff.}$, 3.2.

Difference in type two, 2.3; $P.E._{Diff.}$, 4.8.

TABLE IX

COMPARISON OF SCHOOLS 3-7 ("NON-CRUTCH" SCHOOLS) AND SCHOOLS 1 AND 2 ("CRUTCH" SCHOOLS) WITH RESPECT TO SCORES ON THE FRACTION TEST; CREDIT ALLOWED FOR UNREDUCED ANSWERS

Scores: Credit Allowed for Unreduced Answers	Schools 3-7 ("Non-crutch" Schools)	Schools 1 & 2 ("Crutch" Schools)
15	51	26
14	39	9
13	22	4
12	6	10
11	5	6
10	5	5
9	5	3
8	5	3
7	3	..
6	2	1
5	6	4
4	4	1
3	..	..
2	..	1
1	..	2
0	12	5
Total Cases	165	80
Median Score	14.2	12.8
Q. Same	2.2	2.5
$P.E._{Md.}$	.22	.35
Difference in Medians	1.4	
$P.E._{Diff.}$	.41	
$\frac{\text{Difference}}{P.E._{Diff.}}$	3.41	

"crutch" children was different for these examples from what it was for types one and three. With the second type of example these children not only used a greater number of different faulty procedures but also used them more frequently. At this point, however, the inquiry was blocked; an impasse was reached when an explanation (for the change in faults) was sought in use of the "crutch." It was quite impossible to find anything in the nature of the "crutch" which could account for the discovered facts.

The conclusions, then, regarding the third objection to be raised against teaching the "crutch" were five in number: first, the "crutch" children and "non-crutch" children seemed to be about equally skilled in securing completely correct answers to examples in the addition of proper fractions; second, the "non-crutch" children seemed to be somewhat superior to the "crutch" children in this arithmetical process when unreduced answers otherwise correct were given full credit; third, the "crutch" children experienced their chief difficulty with the type-two examples; fourth, the reason for this difficulty could not in any way be found in the nature of the "crutch" itself; fifth, as a matter of fact all the small variations noted in the ability of the two groups of children to add proper fractions might readily be explained as due to instructional differences other than those associated with whether the "crutch" had or had not been taught the children compared.

Conclusion to Part II. The purpose of Part II was stated to be twofold: to present

TABLE X

COMPARISON OF SCHOOLS 3-7 ("NON-CRUTCH" SCHOOLS) AND SCHOOLS 1 AND 2 ("CRUTCH" SCHOOLS) WITH RESPECT TO RELATIVE SUCCESS ON THREE TYPES OF EXAMPLES IN THE FRACTION TEST

Number Examples Correct	Examples 1-5		Examples 6-9		Examples 10-15	
	Schools 3-7	Schools 1 & 2	Schools 3-7	Schools 1 & 2	Schools 3-7	Schools 1 & 2
6	...	...	...	...	69	37
5	119	55	...	...	38	11
4	25	13	106	47	21	10
3	6	3	22	11	4	4
2	...	2	13	8	7	2
1	4	3	2	1	2	3
0	11	4	22	13	24	13
Cases	165	80	165	80	165	80
Total Examples Correct	717	343	518	218	716	336
Possible Number Correct	825	400	660	320	990	480
Per cent Correct	86.9	85.8	78.5	68.1	72.3	70.0

part of the basic data and to explain the measures by which those data were interpreted. Part II has therefore no special conclusions of its own; the conclusions for Part II are necessarily the same as those for Part I in as much as the two sections of the report simply represent two different methods of looking at the same problem and its solution.

"INTELLIGENCE" AND "PERSEVERATION" RELATED TO SCHOOL ACHIEVEMENT[1]

by
K. H. Rogers,
University of Toronto

This research explores the possibility of examining records of school progress in such a way as to show the dependence of achievement upon certain measurable "factors." Since the development of group tests, many investigations have dealt with the relationship between intelligence and school progress.[2] In the main, these studies have not gone very far beyond the statement of results, or the postulation of special abilities, interests and the like. Recently, however, Brigham contended that if techniques could be devised for isolating some of the factors other than intelligence which are necessary to explain achievement scores, an important field of research would be opened; for the understanding of the nature of these other factors (called by him "societal factors") might contribute to our knowledge of important training differences which play a part in determining lines of interest and success (3).

In recent years, a number of "factors" other than intelligence have been statistically demonstrated by Spearman (12), Thurstone (14), and Kelley (6). Most of the discussion concerning these "factors", however, has been centered upon the validity of the techniques involved, or upon the gross relationships between the "factors" and other major variables. Of these "factors", the one symbolized by Spearman as "p" ("perseveration") has received most attention. But the examination of "p" has never been carried in the direction of its possible connection with the disparity between "g" and school achievement.

Accordingly, our problem may be stated in terms of: (1) an attempt to see if the importance of "intelligence" as reflected in measures of school progress varies in widely different educational settings; and (2) in the event of the correlation being comparatively low under some conditions, does "p" play any part in explaining the fact of this low correlation?

The first study, therefore, dealt with the achievement of subnormal patients in academic and manual fields at the Ontario Hospital, Orillia. By adopting certain of the techniques formulated by Spearman (12), an attempt was made to clarify the dependence of achievement upon general intelligence. In the second part of the research, similar results for public school children of approximately the same mental age were obtained. A comparison was thereby made possible between the results with subnormal and with normal subjects. Since this comparison indicated the dependence of public school achievement on factors other than "g", particularly on "perseveration", the third setting explored the relationship between "perseveration" scores and school progress.

II. The Research

A. Examination of Subnormal Children

(i) Procedure

The first experiment was conducted at the Ontario Hospital, Orillia.

Records in all school subjects were available for a group of fifty-one cases (21 girls, 30 boys). The age, mental age, and I.Q. ranges of these children are indicated

1. An abstract of a thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy in the University of Toronto.

The writer wishes to express his thanks to Dr. W. Line, University of Toronto, for assistance respecting points of method and interpretation, and for his generosity with time and effort in checking the statistical data and techniques employed in this study; and to Dr. S. Horne, Superintendent of the Ontario Hospital, Orillia; Principals S. Richardson and J. Vickery of Regal Road and Balmy Beach Public Schools for permission to use subjects from their schools; and to the members of the Department of Psychology, University of Toronto, for their interest throughout.

2. Many Journal references could, of course, be given here. An indication of these is to be found in the bibliographies given in Pintner (10) and St. John (13).

as follows:

	C.A.	M.A.	I.Q.
Average	14 yrs. 1 mo.	9 yrs. 10 mo.	61.9
Range	10 yrs. 1 mo. 16 yrs. 2 mo.	7 yrs. 6 mo. 11 yrs. 5 mo.	50-70

It will be seen that the cases were confined to the "moron" group, whose mental-age level corresponds approximately to that between the senior first and junior third grades in the Ontario Public Schools.

The subject-scores available included those for arithmetic, reading, spelling, composition, writing, social science, physical training for all cases, manual training for the boys and sewing for the girls. Comparable measures were obtained in all these subjects.

(11) Results

The intercorrelations of these results are shown in Table I.

Two things will be noted immediately concerning this table. The first is the uniformly high correlation coefficients. The second, is the absence of any significant tetrad-difference in that part of the table involving all cases. The probable error of the tetrad-difference is .028 (8) and the median .03.

TABLE I
CORRELATIONS BETWEEN ACADEMIC AND "MANUAL" SUBJECTS

	A	B	C	D	E	F	G	H	I
A. Arithmetic	-	.89	.83	.90	.83	.85	.78	.82	.86
B. Reading	.89	-	.92	.89	.85	.83	.71	.85	.79
C. Spelling	.83	.92	.-	.90	.81	.86	.70	.86	.70
D. Composition	.90	.89	.90	-	.89	.86	.82	.88	.90
E. Writing	.83	.85	.81	.89	-	.94	.80	.80	.90
F. Social Science	.85	.83	.86	.86	.94	-	.71	.81	.80
G. Physical Training	.78	.71	.70	.82	.80	.71	-	.83	.92
M.A.	.78	.75	.69	.71	.72	.78	.73	.72	.92
H. Manual Training Boys: 30 cases	.82	.85	.86	.88	.80	.81	.83	-	-
I. Sewing Girls: 30 cases	.86	.79	.70	.90	.90	.80	.92	-	-

Mean r = .83
Average p.e. of r = .042
N - 51 cases (except where otherwise stated)

It is impossible, from these data, to determine accurately the relationship between the "manual" subjects (manual training for the boys and sewing for the girls), and the other variables. For our immediate purpose, however, it appears to be sufficient to consider the following facts. The mean of the coefficients for the main part of the table (including all cases) is .837. The values in the "manual training" column also give an average of .837; and for "sewing" the mean is .838. There is no evidence, therefore, of any "group factor" in this situation.

The only conclusion possible from the above seems to be that the measures of achievement here employed show a high dependence upon "g", but do not reflect the influence of any other major factor. In other words, not only does this particular institutional programme result in a high relationship between "intelligence" and achievement--a fact that appears to be of considerable educational importance--but also any difference between achievements in verbal and non-verbal fields fail to make themselves evident. Especially noteworthy is the latter finding when we consider the fact that "manual" subjects usually correlate to comparatively slight degrees with academic subjects in the case of normal children in the public schools. The greater degree of standardization in the hospital generally, and more especially with respect to manual work, is probably an important factor in making the correlations high. But this only emphasizes again the possible influence of well-directed education in endeavoring to maintain a high correlation between "intelligence" and achievement.

B. Examination of Public School Children

1. Intelligence and Achievement

The next step was to investigate the same relationship with children in the public school.

(1) Procedure

In order that the mental age range might be somewhat comparable with that of the earlier group of subjects, two classes in the junior III grade were chosen. The pupils of these classes were representative

of all the junior III pupils in the school. This was possible because of the approximate equalization of the groups in any one grade at Regal Road School, Toronto, resulting from the grading policy of the Toronto Research Division of the Canadian National Committee for Mental Hygiene. The pertinent data concerning this group are summarized as follows:

Age Range: 9 yrs., 4 mo. to 13 yrs., 3 mo.
I.Q. Range: 73 to 143.
No. of Cases: 80.

The achievement records in the main school subjects were obtained from the teachers of these classes. Since the results of correlating the total school standing at different two months' periods indicated satisfactorily high reliability (.89), the existing measures were deemed to be adequate to our purpose. For correlation purposes, the marks of the two classrooms were distributed separately and converted into sigma deviation units. "Intelligence" measures were obtained by means of the Pintner Group Test.

(ii) Results

Table II shows the intercorrelations obtained. It will be noted that the correlations are here much lower than was the case with the group of subnormals (see Table I). They vary from .87 to .07 with an average of .452. With the subnormals, the mean of the correlations was .837 with a range of .92 to .70. While the school subjects involved in these tables vary somewhat, the difference in the results obtained is none the less striking. A similar difference has been found by other investigators, and will be discussed later.

A second characteristic of this table is that a number of the tetrads indicate the presence of group factors. For example, one of the tetrad-differences obtained with variables, F, D, H, E is .262 with the probable error of .05. Other examples need not be quoted at this point, since the distribution of tetrad-differences is given later.

TABLE II
TABLE OF INTERCORRELATIONS

	A	B	C	D	E	F	G	H	I	J
A. Total Standing	-	.87	.83	.69	.72	.62	.67	.53	.61	.40
B. Composition	.87	-	.64	.65	.56	.43	.52	.46	.50	.39
C. History	.83	.64	-	.39	.58	.57	.52	.43	.58	.25
D. Spell.	.69	.65	.39	-	.44	.22	.29	.62	.19	.37
E. Lit.	.72	.56	.58	.44	-	.55	.34	.36	.53	.15
F. Pintner Intelligence Test	.62	.43	.57	.22	.55	-	.42	.35	.44	.13
G. Arith.	.67	.52	.52	.29	.34	.42	-	.20	.39	.22
H. Read.	.53	.46	.43	.62	.36	.35	.20	-	.07	.19
I. Geog.	.61	.50	.58	.19	.53	.44	.39	.07	-	.12
J. Writing	.40	.39	.25	.37	.15	.13	.22	.19	.12	-

2. Factors other than "Intelligence" reflected in achievement scores

With the public school children, the low correlations obtained suggest that the standards represented by school marks depend not only upon "g" but also upon some group factor or factors. Accordingly, consideration was given to the nature of these latter.

(i) Procedure

In addition to the variable already discussed (Table II), scores were obtained by means of the following tests:

(1) "P" tests.- A battery of 4 tests was used. They were simple pencil-and-paper tests based on the principle of "perseveration" as investigated by Lankes (7), Bernstein (2), Pinard (11) and others. (For convenience, in Table III, the scores are reversed to give a larger number of positive correlations, so that they more truly represent "c" scores.)

(2) "Alertness."- The tests reported elsewhere (8) were administered in order to obtain speed records with very easy intellectual tasks.

(3) "Speed of Execution."- The number of letters, digits, etc., written in short-time periods, under conditions of speed, served as the main principle here (8).

(4) "Fluency."- This test depended on the number of different ideas that could be written in five minutes, according to the techniques employed by Hargreaves (5). It was included as one of the variables in an attempt to see how far "fluency" and "perseveration" (inertia) were related as psychological manifestations.

These test batteries were chosen mainly because they represented material with which group factors had already been isolated by previous investigators. With each of these batteries, fairly high reliability (self-correlation above .85) was obtained, with the exception of the "p" tests, where the material was not so easily administered. But even though the self-correlation here was only .5, the use of the results of these tests in the manner indicated below seems to be justified.

The next step was to calculate inter-correlations throughout the whole table of variables. This table was then considered in the light of the tetrad-difference (Two Factor) method of Spearman (12).

(ii) Results

The additional correlations (i.e., additional to those given in Table II) are presented in Table III. The distribution of tetrad-differences for the correlations of both tables combined is given in Table IV.

TABLE III
TABLE OF INTERCORRELATIONS

	K	L	M	N
A. Total Standing	.13	.09	-.23	.20
B. Composition	.23	.11	-.14	.20
C. History	.12	.11	.09	.10
D. Spelling	.28	.12	-.16	.31
E. Literature	-.05	.00	-.04	-.04
F. Intelligence	.09	.02	-.10	-.01
G. Arithmetic	-.11	.07	-.37	.08
H. Reading	.13	.23	-.10	.32
I. Geography	.01	-.09	.04	-.16
J. Writing	.37	.31	.05	.33
K. Fluency	-	.11	.41	.23
L. Alertness	.11	-	.13	.17
M. Speed of Execution	.41	.13	-	-.25
N. Perseveration	.23	.17	-.25	-

The relations of reading, spelling and writing with "c" are interesting, especially when contrasted with the apparent correlation of literature and geography with "perseveration." Further detailed analysis provided supplementary and corroborative data for these findings. As a contrast to geography, history was similarly analyzed, but showed no such clear-cut influence of "perseveration."

TABLE IV
THE DISTRIBUTION OF TETRAD-DIFFERENCES FOR TABLES II AND III

Tetrad-differences	Frequency
.40 to .45	1
.35 - .40	3
.30 - .35	10
.25 - .30	38
.20 - .25	125
.15 - .20	247
.10 - .15	492
.05 - .10	774
0 - .05	1313
	3003

Median tetrad-difference = .062
P.E. of tetrad-difference = ±.038

3. Some qualitative considerations

A further lead to interpretation seems to be offered by studying the relationship between factors such as "perseveration" and "case history" material. Of the children used in this investigation, twenty-five had been examined by the Research Clinic in connection with behavior and educational problems. Ten of these cases obtained "perseveration" scores that placed them in the low 25 per cent of the "perseveration" distribution, five were high "perseverators" and ten were within the middle fifty per cent. The first group were reported to the Clinic as being: (a) badly behaved in school, restless, inattentive, unreliable, careless, given to minor forms of misdemeanour (3 cases); (b) involved in sex delinquencies in situations where the major responsibility was deemed to rest on other children present (2 cases); (c) shy, sensitive, nervous (1 case); (d) bright, attractive, a trifle spoiled, restless (1 case); (e) sensitive, talkative, truant, ready with plausible reasons for his behavior (1 case);

(f) inattentive, weak in geography, careless (1 case). The five high "perseverators" were described as deceitful, lacking in enthusiasm, unwilling to do particular work (1 case); very weak in reading and spelling (2 cases); the "worst" child in the room (1 case); a sex delinquent, -- the "worst" boy in the neighborhood (1 case). In contrast, the ten children of the normal "perseveration" group were reported as being dull (2 cases); for routine examination in connection with the history of a brother or sister (3 cases); health problems (2 cases) the "best" child in the class (2 cases); and only one as a "nervous" case. While no conclusions can be drawn from such scanty data, it is interesting to note the harmony between the picture here represented and the findings of Pinard (11); for this investigator found that "difficult" cases tended to show high or low "perseveration", the high "perseverators" being the more persistent in their delinquency.

Much elaboration of this pattern of research is necessary before we can interpret any of the factors further. The number of cases is small; and observations in different educational and social settings would have to be made before any definite interpretation could be offered.

C. Relationship between School Achievement and "Perseveration"

The data presented in the last section indicated that of the factors other than "g" reflected in school marks, "perseveration" seemed to be the most significant. A further study was therefore made with a view to seeing if the relationship between "perseveration" and achievement could be more definitely determined.

(i) Procedure

For this purpose another school was chosen. These children were first examined by the Pintner Intelligence Test. The age, mental age ranges, and number of cases in each grade are shown in Table V.

Then "perseveration" scores were secured, using tests similar to those employed above. These were administered at approximately the same time as were the "intelligence" tests. The school standing was determined by the marks given by the teachers on the examinations of the month immediately preceding.

TABLE V
THE SUBJECTS ACCORDING TO GRADES

Grade	No. of Cases	C.A. Range	M.A. Range
Sr. IV	38	12.7*- 16.6	12.9 - 16.7
Jr. IV	36	10.9 - 14.8	11.6 - 15.5
Jr. III	28	9.2 - 13.7	10.1 - 14.1
Jr. III	24	9.2 - 13.4	10.9 - 13.8
Jr. III	24	9.0 - 13.7	9.10- 13.7
II	24	8.2 - 11.6	9.6 - 11.9
II	26	7.6 - 11.3	9.4 - 11.10
II	20	8.0 - 11.6	9.1 - 11.9
Total	220	8.0 - 16.6	9.1 - 16.7

*i.e., 12 years, 7 months.

The reliability of the "p" tests is to some extent indicated by the correlations given in Table VI.

TABLE VI
INTERCORRELATIONS OF "p" TESTS

	(1)	(2)	(3)
(1)	-	.26 (.04)	.32 (.04)
(2)	.26	-	.23 (.04)
(3)	.32	.23	-

Number of cases: 220
Average correlation: .27
Average probable error: .041

By taking the first test against the second and third, a correlation of .45 was obtained. Comparing these values with similar ones reported in the literature (Pinard, Jones, etc.), they suggest that results obtained by a group examination show sufficient reliability to justify their use in the way here employed.

(ii) Results

The "perseveration" scores were compared with age, mental age, and I.Q. The tables of results showed no significant trend of variation for "perseveration" as here measured in relation to any of these three variables.

An interesting relationship appeared, however, when the disparities between "intelligence" and achievement of the high "perseveratives", who were below the average

of the class in "intelligence", were compared with the low "perseveratives", who were similarly below average in "intelligence." The term "disparities", as here used, can best be understood by referring to the way in which they were determined. For each classroom, the "intelligence" test scores were distributed separately and the point scores were translated into deviation (sigma) units. The same procedure was adopted in connection with the achievement marks. The difference between the respective deviation scores for any one individual constituted his measure of disparity. Where the achievement rating was higher than that for "intelligence", the disparity was called "positive" (+); in other cases it was called negative (-). "High" and "low perseveratives" refer to those cases who deviated more than one sigma in the "perseveration" distribution, that is, the positive and negative extremes.

Respecting the above-mentioned relationship, it was found that whenever "intelligence" is below the average, the tendency is for disparities to be positive irrespective of the "perseveration" rating. The results for high and low "perseveratives" with more than average "intelligence" were distributed, and in contrast with the foregoing, it appears that there is a preponderance of negative disparities among the children who score above the average in "intelligence", a total of 13 negative and 3 positive for the high "perseveratives", and 15 negative and 6 positive for the low "perseveratives."

When the medium positive "perseveratives" are combined with the high "perseveratives", and the medium low "perseveratives" with the low "perseveratives", the general indication is that the inclusion of cases who do not deviate markedly from the mean in "perseveration" reduces the tendency above mentioned, without, however, causing it to disappear. In other words, while it seems characteristic of children who are below the average of their class in "intelligence" to manifest positive disparity (and contrariwise for those above the average in "intelligence"), this tendency is accentuated with both the high and low "perseveratives."

In order to obtain corroborative data on this point, the disparities of the high and low "perseveratives" throughout the school were compared with those pertaining to the medium group. In the first place, the children in each classroom were divided into four equal groups (marked off by the quartiles) on the basis of their "perseveration" scores, and then further classified in accordance with their "intelligence" ratings as to whether they were above or below the average. Again, as above, there was found a difference with respect to "perseveration." The tendency for the children of more than average "intelligence" to show negative disparity and for those of less than average "intelligence" to show positive disparity is much greater with the extremes of the "perseveration" distribution than with the others.

Further corroborative evidence was demonstrated when the average amount of disparity displayed by these various "perseveration" groups was determined. Table VII shows that in the fourth grade the high "perseveratives" who are above the average in "intelligence" show an average negative disparity of .80; while the high "perseveratives" who

TABLE VII
THE AVERAGE AMOUNT OF DISPARITY (ACHIEVEMENT FROM "g") DISPLAYED BY THE "PERSEVERATION" GROUPS

	High "p"		Low "p"		Middle 50%	
	+"g"	-"g"	+"g"	-"g"	+"g"	-"g"
4th	-.80	.81	-.96	.88	-.28	.60
3rd	-.81	.50	-.44	.61	-.66	.84
2nd	-1.18	1.16	-.09	.90	-.53	.48

are below the average "intelligence" show an average positive disparity of .81. The other values in the table, read similarly, again draw attention to the fact that the disparity values tend to be higher with the high and low "perseveratives" than with the middle 50 per cent. This is especially true in connection with the high "perseveratives" of the second grade and with the low "perseveratives" who are below the average "intelligence" in this grade.

In order to make sure of the fact that these figures are not seriously influenced by the inclusion in the medium "perseveration" group of both the medium positive and medium negative cases, it was decided to obtain corroborative correlational results. If the above contention is true, those cases who are within this middle 50 per cent group of the "perseveration" distribution should show greater correlation between "intelligence" and achievement than is the case with extremes (high and low "perseveratives"). The actual values obtained are .42 and .25 respectively. In other words, while there is a tendency for all children who are below the average "intelligence" of their class to show positive disparity, and for those who are above the average "intelligence" to show negative disparity, this tendency is enhanced in the case of both the high and low "perseveratives."

III. Summary and Discussion of Results

1. The first study indicated the relatively high importance of "intelligence" in the academic and "manual" achievements of subnormal children. There seemed to be no definite manifestation of "group factors" in this setting.

2. With a fairly comparable group of public school children, the relationship between "intelligence" and achievement in school subjects seemed to be much lower than in the former case, and "factors" other than "intelligence" appeared to underly some of the variations in achievement in different subjects. "Perseveration", as measured by the tests employed, seemed to be particularly important. Some relationship between this factor and behavior tendencies of the children is suggested.

3. The third study investigated this factor "perseveration" more extensively by examining 220 children of the public school. There seemed to be no relationship between "perseveration" and age, mental age, and I.Q. However, a relationship appeared to be indicated when "perseveration" scores were compared with "disparities" in achievement with respect to "intelligence." While there was a tendency, common to all grades investigated, for the children of more than average "intelligence" for their grade to show negative "disparity", and for those with lower "intelligence" to show positive "disparity", this tendency was increased in the case of the upper and lower 25 per cent of the "perseveration" distribution.

The difference between the achievements of subnormal and normal children with respect to their dependence on "intelligence" has been noted previously. For example, Abelson (1) applied the same tests to both classes, and for normal children obtained a mean correlation of .466, while for the defective children the value was .782. Spearman notes a similar condition present in comparing young and older children (12, p. 218), and in comparing children with adults (12, p. 219). Spearman's interpretation of this tendency is well known, (12, p. 220). He uses an analogy, the law of "diminishing returns", and states that the influence of any general factor in differentiating the achievement of different individuals grows less as the factor itself increases. Here, while this fact may be adequate to explain our results in part, we should note, among other possible interpretations, the following: (a) the range of talent is somewhat greater with the mental deficients here represented, than with the public school group employed in the second part of the study. This and other factors may therefore have influenced the reliability of the measures concerned, in favor of those pertaining to the hospital situation. Information available for our data, however, does not suggest that this is more than a minor factor in the situation. (b) The hospital setting may have contributed a "halo" effect, in that the pupils are being marked by teachers who associate with them continuously. But this is even less true of the hospital group than in the public school, since in the latter one teacher only gave the scores that were used, while in the former several teachers were involved. Further, the marks available with the subnormals were in many cases obtained from "objective" tests constructed by the principal on the basis of the curricular demands. On the other hand, the careful standardization of the curriculum at Orillia, and the uniformity of the institutional background,

may contribute considerably to this phenomenon.

Concerning the part played by "perseveration" in school achievement, any interpretation at this stage must of necessity be very tentative. But the writer draws attention to its possible relationship to school procedures as suggested by the results of the second study. The subjects favored by the "perseverative tendency" (notably, literature and geography) would appear to give more scope for a contemplative type of thinking, and this probably offers a lead to the interpretation of "perseveration." If this is the case, an interesting point arises respecting the results obtained with history, where "perseveration" seemed to be of no significance, which suggests the possibility that a greater emphasis is put on isolated facts in this subject.

Whatever the interpretation of "perseveration" may be, the origin of this factor appears to be traceable at least to the early years, since it showed no significant increase with chronological age in our study. Since the incidence of high or low "perseveration" has been found by Pinard to be unusually high with problem cases--a fact that is to some minor extent corroborated by our results--the nature of early home training may have a great deal to do with the presence of this factor. Further, when this factor is present in either extreme, it would appear to have some influence on the adjustment of the child to the demands of the classroom, as the results of disparity cases tend to indicate.

IV. Conclusions

(1) When the achievements of subnormal children are compared with those of normal children in the fields represented by school subjects, the former tend to show a much higher relationship with "intelligence" than do the latter.

(2) The fact stated in conclusion (1) would seem to indicate a much more carefully directed educational situation in the case of the subnormals in this particular setting.

(3) It appears to be difficult to isolate and interpret the significance of measurable factors other than "g" with subnormal children in a setting such as the one chosen.

(4) Psychological factors other than "g" are definitely reflected in the subject-achievements of normal school children.

(5) Of the factors referred to in conclusion (4), "perseveration", at the present time, appears to warrant the most emphasis for further study.

(6) Within the age-range of the children studied (7 to 16 years), no evidence of increasing "perseveration" with age or with mental age was demonstrated.

(7) "Perseveration" appears to be related to the tendency herein shown, for children of more than average "intelligence" to display negative disparity between "intelligence" and school achievement; and with the contrary tendency, for those of less than average "intelligence" to show positive disparity. High and low "perseverators" display these tendencies to a greater degree than do those children who are within the medium ranges of the " perseveration" distribution.

BIBLIOGRAPHY

1. Abelson, A. R. "The Measurement of the Mental Ability of 'Backward' Children", British Journal of Psychology, IV (1911), pp. 268-314.
2. Bernstein, E. "Quickness and Intelligence", British Journal of Psychology, Monograph Supplement, III, No. 7, (1924), pp. vi + 55.
3. Brigham, C. C. A Study of Error. New York: College Entrance Examination Board, (1932).
4. Jones, L. Wynn. "Individual Differences in Mental Inertia", Journal of the National Institute of Industrial Psychology, IV (1929), pp. 282-94.
5. Hargreaves, H. L. "The 'Faculty' of Imagination", British Journal of Psychology, Monograph Supplement, No. 10, (1927), p. 74.
6. Kelley, T. L. Crossroads in the Mind of Man (Stanford, California: Stanford University Press, 1928), p. 238.
7. Lankes, W. "Perseveration", British Journal of Psychology, VII (1915), pp. 387-419.
8. Line, W., and Kaplan, E. "The Existence, Measurement and Significance of a Speed

Factor in the Abilities of Public School Children", Journal of Experimental Education, I (September, 1932), pp. 1-8.

9. Research in a Hospital for Subnormal, Ontario Hospital, Orillia, Publications, Vol. I, No. 1, 1929; Vol. I, No. 2, 1931.
10. Pintner, R. Intelligence Testing (New York: Henry Holt and Company, 1930) p. 555.
11. Pinard, J. W. "Tests of Perseveration, I: Their Relation to Character", British Journal of Psychology, XXIII (1932), pp. 5-19; "Tests of Perseveration, II: Their Relation to Certain Psychopathic Conditions and to Introversion", British Journal of Psychology, XXIII (1932), pp. 114-26.
12. Spearman, C. The Abilities of Man (New York: Macmillan Company, 1927), pp. 415 + xxxiii.
13. St. John, C. W. Educational Achievement in Relation to Intelligence (Cambridge, Mass.: Harvard University Press, 1930), p. 200.
14. Thurstone, L. L. "Multiple Factor Analysis", Psychological Review, XXXVIII (1931), pp. 406-27.

THE EFFECT OF PRAISE, REPROOF AND EXERCISE UPON MUSCULAR STEADINESS

by
Merrill T. Eaton
Indiana University

The use of praise or reproof to secure increased efficiency in individuals is perhaps more frequently resorted to than any of the other more common incentives. In education, in business, in athletics, in the home, and elsewhere the use of praise and reproof is an everyday occurrence. A clear-cut and satisfactory explanation of just what effect is produced in the individual when he is praised or reproved has, however, never been given. A survey of the investigations dealing with this problem, as well as the opinions of various writers on the subject, would lead one to believe that praise or reproof as an incentive affects the behavior of the individual through certain changes produced in his emotional state.

An attempt at an analysis of this problem leads one into many diverse fields and through many different types of investigation. A review of some of the typical studies in these fields shows the complexity of the problem. Gilchrist (7) investigated the extent to which praise or reproof affects a student's work. In this study fifty students enrolled in a class in educational psychology, after they had taken an English test, were divided into two groups without any consideration of the scores they had made. The members of Group I were told they had done very poorly on the test, and the members of Group II were informed that they had done exceptionally well. Both groups were then requested to take the test again. The group that was praised improved its score 79 per cent; the score of the reproved group was lowered 6 per cent. The investigator attributed these results to the attitude set up by his words of praise and reproof. Hurlock (10), testing the effect of praise and reproof, used school children as subjects. Intelligence tests were given the children without the use of any specific incentives, and on the basis of their scores the subjects were divided into three groups having identical averages on the intelligence test. The three groups were then required to take the test again. For one group nothing was said about its performance on the first test; another group was praised for its high scores and neat work; the third group was reproved by being informed that its work was poor and was told that it should be ashamed to make such a low score. All three groups made gains when the test was repeated. The gains for the groups which were praised and reproved were approximately equal. In the control group, 52 per cent of the children improved; in the praised group, 79 per cent; and in the reproved group, 80 per cent. Results secured by the same investigator (9) in testing 106 school children from the fourth and sixth grades showed that praise was a valuable incentive for setting up a motivating condition within the learners which enabled them to do more and better work. Somewhat similar studies by Kirby (12), Gates (6), and Chapman (4) have confirmed these results.

Book (3), in commenting on the effect produced by praise or reproof upon a pupil's reactions and ability to learn, says: "We may conclude that praise by a teacher sets up a favorable motivating condition in the learner that enables him to do more and better work." Bird (2) contends that: "As a form of social recognition, praise acts upon its recipient to arouse attitudes favorable to the project in hand; consequently, it improves the learning process." Kitson (13), investigating the work of forty journeymen printers, found that when a

bonus was given for work completed in excess of the average amount turned out daily, the average daily output for the group was improved 78 per cent.

By far the most extensive studies dealing with the comparative merit of learning by reward and learning by punishment have been made at the Institute of Educational Research of Teachers College, Columbia University, under the direction of Thorndike (18). In interpreting these investigations, Thorndike says: "Other things being equal, an announcement of 'right' strengthens the connection which it follows and belongs to much more than the announcement of 'wrong' weakens the one which it follows and belongs to." Punishment, however, he believes is not always futile, but the results of these studies should increase confidence in positive rather than negative learning and teaching.

Griffith (8), in discussing the value of praise and reproof as incentives in athletics, states that he considers praise a more dangerous weapon than reproof, thereby indicating that he believes the greater emotional stimulation resulting from praise is detrimental to the best performance of the athlete. Berry (1) says: "The fact that athletics possess such a high emotional content is one of the greatest problems in coaching." Because of the high emotional setting in which athletes perform, care should be practiced in the use of praise and reproof. He states further: "As yet no one has learned the technique of this (emotional) phase of coaching."

The investigations reviewed and the points of view enumerated above would indicate that really to measure the effect of praise and reproof some measures of the differences in emotional status of the individual should be attempted. This problem has been attacked from many different angles, using various testing techniques. The galvanometer is an instrument used in one of the more complicated techniques. Symonds (17), however, considers it "a well-established method for detecting the presence of emotion." Landis and De Wick (14) have made a critical review and summary of the literature bearing on the psychogalvanic reflex.

None of the investigations mentioned above, however, is closely related either in purpose or technique to the present study, but all illustrate the many diverse points of attack. Skaggs' (16) study was somewhat similar in technique to the present investigation in that he used a Whipple Steadiness Tester to measure involuntary movements of the hand and arm. His stimulus situations, however, were of the "startle" type, and praise and reproof were not involved. Kellogg (11) found that both the "exercise" and the "shock stimuli" which he employed affected muscular steadiness, causing the subjects to register more contacts per test. He also found that the number of contacts was increased more after moderate physical exercise than by the emotional situations which he produced by the "shock stimulus."

PROBLEM

As a review of the studies in the field clearly shows considerable uncertainty exists in regard to the effects which praise, reproof, or exercise actually produces upon the performance of the individual. Many of the testing techniques previously used have necessitated the use of very complex and elaborate apparatus or testing conditions not fully controlled. This investigation, therefore, was undertaken to measure by a very simple apparatus the changes in muscular steadiness resulting from praise, reproof, or exercise. The experimental conditions were arranged so that all testing was done in the normal, natural learning environment of the subjects, and a testing procedure was followed which would tend to show definitely the effects of the various stimulus situations that were arranged in the present experiment. The chief aim of this study was to determine whether praise, reproof, or exercise affected the muscular steadiness of men and women, and if so, how and to what extent.

EXPERIMENTAL CONDITIONS

Apparatus. The apparatus selected to measure steadiness was a plate-and-stylus tester of the Whipple type. The standard

apparatus, as supplied by Stoelting, was modified somewhat to adapt it better to the experimental conditions under which the tests were given. These modifications previously made by Kellogg (11) were as follows: "(1) A sliding metal cover was built in front of the contact plate to permit the exposure of only one hold at a time. This precluded the accidental insertion of the stylus into the wrong hole. (2) A 6-volt electric lamp was connected through a relay in the plate-stylus so that it flashed whenever contacts were made. This light, mounted almost directly in S's line of vision upon the upper edge of the contact plate, informed him when the stylus touched it and prevented him from falling unintentionally into the error of resting the stylus upon the circumference of the hole and so obtaining a spuriously high steadiness score"; (3) The contact plate was mounted on a tripod, the tripod resting on a small table about thirty inches high. With this arrangement the contact plate could be conveniently raised or lowered depending upon the height of the student. Since this experiment was conducted under absolutely normal or natural learning conditions, the apparatus was located, in case of the men, in the men's gymnasium just outside the boundary line of the basketball court, opposite the center circle. In case of the women testing was done in the physical education building for women, the apparatus being located approximately opposite the center of the pool and as near the edge of the pool as was possible without water being splashed on the equipment. (4) All steadiness records were made with the subject standing. The contact plate was adjusted to the shoulder height of each subject who stood at a distance from the apparatus equivalent to his arm's length.

Subjects and Experimental Procedure. Forty-five persons, twenty-three men and twenty-two women, were used as subjects. All the men were freshmen students at Indiana University and were candidates for the freshmen basketball squad. The testing was done during the month of March (1933) after the men acting as subjects had been participating in basketball practice all season. The women subjects were members of a swimming class; six were freshmen, eleven were sophomores, and five were juniors. The women subjects were tested during the month of May (1933) after they had been participating in swimming twice weekly since the beginning of the second semester, February 1, 1933.

None of the subjects had any knowledge of the purpose of the experiment before it was undertaken or during the course of the experiment. They were informed that the purpose of the tests was to measure steadiness under various learning conditions. The testing period being near the close of the basketball season, the men were motivated to do their best, since those ranking highest among the freshmen candidates were to be awarded "numerals." From remarks made by these men subjects, the experimenter feels sure that they took the experiment seriously because they believed that the results of the study would be considered in making the "numeral" awards. The experimenter was careful neither to deny nor to confirm this belief.

The women subjects were marked on the amount of improvement they made in swimming during the semester, and since they also thought the results of the experiment would be considered in making up their final marks, the experimenter feels confident that they too were motivated to do their best.

Each subject was given practice trials until he was thoroughly familiar with the operation of the equipment. No attempt was made to screen from the subjects the electric counter for recording the contacts made, and all the men and women asked each time for their "score" (i.e., the number of contacts) made during the test period. In fact, they seemed to take special pleasure in knowing that they were able to make a good steadiness "score."

Preceding the test period, neither the experimenter nor his assistants[1] talked to the subjects. They came directly from the basketball floor or the swimming pool and were tested immediately. Following the test, however, the experimenter freely discussed the "score" with the subject. Their introspections and comments were, however, not recorded, a fact that probably constitutes a weakness in the testing technique.

1. Walter Marks, freshman basketball coach, and Miss Erna Van Valer assisted in the testing.

Preliminary tests on several basketball candidates and swimming class students whose results are not included in the present study indicated that a test period of ten seconds would be the most desirable. Kellogg (11), using the same apparatus and a similar testing technique, found that if the test period did not exceed twelve seconds, with a short rest period intervening between tests, fatigue could be almost entirely eliminated.

The procedure followed in making the test was for the experimenter to adjust the contact plate to the shoulder height of the subject, hand the subject the stylus, have him place it in the hole in the contact-plate (the stylus was held in an horizontal plane and perpendicular to the face of the metal sheet) and then on the command "ready", the operator pressed the key that connected the stylus and contact plate with the electric counter, which recorded the number of contacts and simultaneously started the stop watch. When the ten seconds were up, the operator released the key, which broke the connection, thus avoiding any extra contacts after the ten second test period. By following this procedure, only the net number of contacts in the specified ten second testing period was recorded.

The subjects were tested in six different types of stimulus situations. The test "series" for the men consisted of ten tests for a norm. Five tests in all other situations were given, from which the average score for each of the five types of stimulus-situations was computed. In the case of the women, five tests constituted a series for the norm and each of other five "series" of tests consisted of three tests. Not more than one test was given in any test series on the same day.[1] The same subjects, however, might be tested twice on the same day; i.e., one test in the norm series and one test in another of the test series.

STIMULUS SITUATIONS

The stimulus situations that could be used in an experiment of this nature were determined through conference with the freshman basketball coach and the swimming instructor. From this list of possible stimulus situations, the choice was narrowed down to six. The six selected were of such a nature that no artificial condition need be established. Neither would the actual testing interfere with the work of the coach or swimming instructor. The situations selected are the following:

1. Norm stimulus. The men subjects reported at the gymnasium at 2:30 P.M. dressed for basketball practice and then reported to the experimenter. They were asked to sit and rest for five minutes to offset any fatigue or emotional disturbance that walking to the gym and dressing might have aroused. They were then given a ten second period test. A similar procedure was followed for the women, except that they reported at the physical education building for women at 10:00 A.M. dressed for swimming.

2. First exercise stimulus. The basketball coach's practice program consisted of having the men shoot baskets for the first thirty minutes. An assistant kept an accurate record of the time the subjects began practicing, and after twenty minutes of basket shooting, the subject was called to take a test. In the case of the women, the assistant recorded the time the subject entered the water and after practicing a simple, overhand stroke with flutter kick for ten minutes, she was called to take a test.

A difference of practice time between the men basketball subjects and the women swimming subjects necessitated a difference in the intervening periods between the tests for the different stimulus situations.

The practice period for the men subjects was two and one-half hours and for the women forty minutes. The intervening time, therefore, between stimulus situations was considerably shorter for the women. The difference in time of the periods, however, is probably justified by the fact that the men had been participating in basketball for a considerably longer period of time and were, therefore, in much better physical condition than the women for the type of exercise engaged in before the tests were given.

1. The testing for women was not begun until late in the second semester and for that reason the number of readings in each test series was reduced so that there would be sufficient time to complete the testing before the subjects began preparing for final examinations.

3. Second exercise stimulus situation. The third stimulus situation was similar to the first, except that the practice period was longer. For the men the second exercise stimulus consisted of 65 minutes of practice, including basket shooting, fundamentals, plays, and fifteen minutes of scrimmage; the fifteen minutes of scrimmage, which was closely observed by the coach, immediately preceded the testing. In case of the women, the second exercise period consisted of twenty minutes of practice on a simple overhand stroke. Five minutes of this practice that just preceded the test consisted of trial strokes under the observation of the instructor.

4. Praise stimulus situation. The conditions for the stimulus situation where the praise stimulus was given were exactly the same as those of the second exercise stimulus, except that in case of the men the coach called the subject to the side of the court, put his arm on his shoulder and there, where no one else could hear, praised the man for some phase of his playing. No subject, however, was praised unless he really deserved it, and the praise was not of the "flowery" type, but that generally used by the coach. The commending statements generally used were: "That perfectly timed pivot helped to pave the way for that basket. It was well done."

Not more than thirty seconds intervened between the time the subject was called from scrimmage and his reporting to the experimenter to be tested. In case of the women the instructor called the subject from the pool and praised her for the improvement she was making. The subject was not praised unless she deserved praise, and the commending statements were of the type the instructor generally used, such as: "Your coordination was excellent. You are improving very much."

5. The Reproof stimulus. The conditions for the reproof stimulus situation were exactly the same as those for the praise stimulus situation and the testing procedure was identical, except that for this stimulus situation the subjects were reproved. They were not criticized, however, unless they deserved it. The reproof used by the basketball coach was, however, not of the rough type, but consisted of constructive criticism; e.g., "If you had followed the path of your pass, you would have been in a good position for a follow-up shot. This is an essential point that you must remember." The swimming instructor also used constructive criticism such as: "That was incorrect. The timing of your stroke should be in better rhythm."

6. Final exercise stimulus situation. The last test for the men was given at the end of two and one-half hours practice; for the women after forty minutes of swimming. No emotional situations were involved in this last test except those which might naturally arise from the type of learning and exercise involved. The subjects were neither praised nor reproved on the days that tests were made for this stimulus situation.

RESULTS

The objective data, average number of contacts per test series, constitute the criteria by which the changes in emotional status of the learner following praise, reproof, and exercise are determined. Introspective data would probably have added to the validity of the conclusions, but such data were not recorded because an attempt to secure from the subjects following each test a statement concerning their "feelings" would have informed the subjects of the purpose of the investigation, thereby changing the testing environment from a normal, natural learning situation to an artificial one.

The procedure used, i.e., individually praising or reproving the subjects should also add to the reliability of the data. Sims (15) found that individual motivation was considerably more effective for producing changes in the emotional state of the learner than was group motivation. Add to this the fact that both the praise and the reproof bestowed on the subjects during the investigation were of exactly the same nature as that habitually used by the coach and the swimming instructor, and one may expect that any changes in the average number of contacts for the different test series, following the stimuli employed, would actually measure a difference in the emotional status of the learner. The writer feels

TABLE I

AVERAGE NUMBER OF CONTACTS PER TEN SECOND PERIOD FOR EACH STIMULUS SITUATION

Subjects	1		2		3		4		5		6*	
	M	W	M	W	M	W	M	W	M	W	M	W**
1	27	9	36	14	39	22	50	21	35	27	43	13
2	12	12	16	15	35	12	35	12	15	20	44	11
3	25	20	32	25	36	25	25	17	40	22	37	23
4	14	18	19	25	23	21	57	24	20	22	27	17
5	18	18	26	17	32	19	39	14	26	17	29	20
6	16	19	22	37	21	33	37	22	29	31	38	22
7	20	15	25	11	29	13	46	32	19	25	37	11
8	16	28	17	31	32	28	21	26	15	26	35	39
9	21	28	20	41	33	28	56	24	40	29	37	27
10	18	16	21	22	25	21	24	16	20	20	31	17
11	20	18	24	17	34	30	42	19	36	25	50	32
12	21	21	29	22	30	25	43	24	35	27	55	21
13	31	20	31	20	47	25	59	20	39	17	47	21
14	28	18	32	21	44	24	39	23	30	20	35	23
15	23	27	26	27	26	29	41	18	36	34	31	25
16	17	16	25	21	33	17	49	21	16	23	29	18
17	19	12	25	14	30	17	52	9	36	16	41	14
18	19	14	21	14	31	21	41	19	35	23	31	16
19	35	19	39	26	39	32	44	23	30	30	40	29
20	18	20	23	24	31	27	42	17	38	21	39	22
21	15	10	25	14	27	16	26	19	17	17	32	17
22	19	16	22	18	29	24	39	20	35	25	32	17
23	18		25		27		28		26		20	
Ave.	20	18	25	22	32	23	41	20	29	24	36	21
S.D.	5.76	6.33	6.13	9.68	6.71	6.34	10.50	5.45	11.87	6.50	8.53	7.92

*Numbers refer to the six types of stimulus situations.
**M refers to Men subjects.
W refers to Women subject.

justified in making this statement because he is measuring changes in the physiological condition of the individual.

Individual differences. A study of the average number of contacts made by individual subjects for the six different stimulus situations (Table I) shows that there were marked individual differences in steadiness. In fact, individual differences appeared in the first testing series (norm stimulus) and were apparent throughout the subsequent testing periods. Other investigators in this field have found similar differences, indicating that normally some learners are more steady than others and that different stimulus situations affect individuals in widely varying degrees. The present data also indicate that there are important sex differences in the emotional changes resulting from the same stimuli.

Arranging the individual scores from highest to lowest and plotting the resulting curves (Figure 1) there results a wide range in the number of contacts made by the subjects for the different stimulus situations. The number of contacts made by the men for the norm ranged from 12 to 35, a difference of 23 contacts between the most steady and the least steady individual; the range was almost as great for the women (9-28), a difference of 19 contacts between the highest and the lowest.

For Stimulus 3, the number of contacts made by the men varied from a low of 16 to a high of 39, again showing a difference of 23 contacts between the most steady and the least steady. The most steady woman subject for Stimulus 3 registered 11 contacts as compared with 41 for the least steady individual, making a difference of 30 contacts between

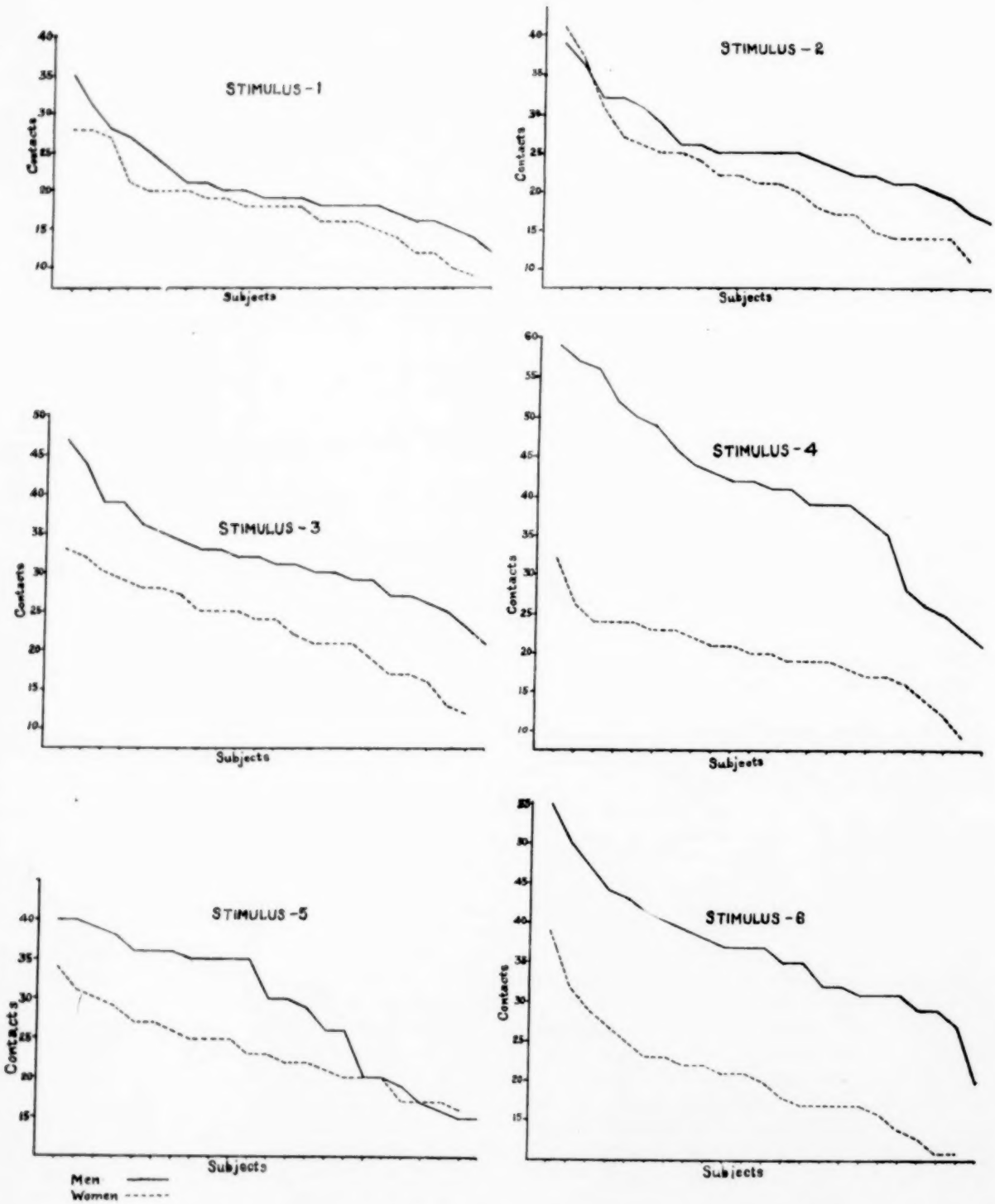


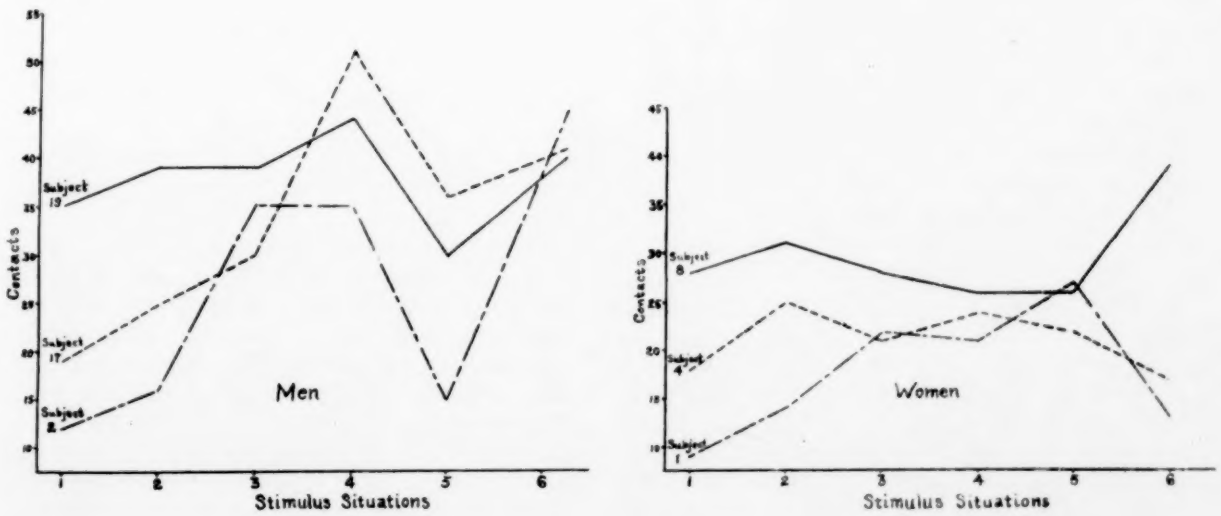
Figure 1. Steadiness Curves for the Different Stimulus Situations.

the highest and the lowest. If one examines the steadiness curves (Figure 1) for the remaining three stimulus situations, it is noted that the conditions described above are, in general, duplicated.

A study of the individual record of each subject (Table I) comparing the number of contacts made for different stimuli shows the varying effect of the stimuli on the emotional status of the subjects. Three subjects, however, have been selected from each sex group as representative of the entire group as a means of making a more critical study of individual differences in steadiness for the different stimulus situations: the subject scoring highest (least steady), the subject scoring lowest (most steady), and the subject making the middle score. In cases where more than one subject made the highest, lowest, or middle score, the first subject appearing in Table I with that score was selected.

The steadiness curves for the three men subjects and the three women subjects selected are shown in Figure 2, the curves on the left representing the performance of the men and the curves on the right showing the performance of the women. These curves portray the individual differences in steadiness resulting from the different stimuli. It can readily be seen that S 19 (M-left) was only mildly affected by a short period of exercise; praise caused only a slight additional decrease in steadiness, but reproof had a remarkably steadying effect by increasing the steadiness to a degree greater than existed for the norm. The long exercise period (stimulus 6) produced little change (5 contacts) over the norm. S 17 (M-left) was affected more by the two exercise stimuli, and praise further decreased the steadiness to a marked extent. Reproof affected S 17 to approximately the same extent as it affected S 19, and the long exercise stimulus produced slightly more effect. S 2 (M-left) decreased only slightly in steadiness for the first exercise stimulus, but the second exercise stimulus caused a pronounced change in steadiness. Praise did not cause any further decrease in steadiness, but reproof affected the emotional status, increasing the steadiness to approximately that of the norm. The long exercise stimulus decreased the steadiness more than that of either S 19 or S 17, as is shown by the curves.

S 8 (W-right) was only mildly affected by Stimuli 2, 3, 4, and 5, steadiness decreasing only slightly after the first exercise stimulus, then gradually increasing in steadiness up to Stimulus 6 where a marked decrease in steadiness was shown. S 4



(W-right) decreased in steadiness for the first exercise stimulus, but steadiness increased for the second exercise period. Praise and reproof produced only a slight change in the emotional status, the curve rising slightly for praise and falling to almost the same extent for reproof.

For Stimulus 6, however, S 4 was almost as steady as for the norm. S 1 (W-right) was affected considerably more for the first two exercise stimuli than was either S 4 or S 18. Praise, however, produced no change in the emotional status, but reproof sent the curve upward higher than for either S's 4, 5, or 6. For Stimulus 6, steadiness increased equal to that recorded for Stimulus 2.

Although none of the individual curves for the other thirty-nine subjects is exactly similar to those presented in Figure 2, some curves showing greater and some less individual variation in change in emotional status from one stimulus to another, a further study of these individual curves does not add materially to our knowledge of how individual learners are affected by the different stimuli presented in this study.

A further study of the effect produced on the learners by the five different stimulus situations, however, does indicate that the emotional status of men and women is affected and to different extents by the same stimulus. A study of the six sample cases (Figure 2) shows that steadiness in both the men and the women was decreased (in different amounts) by a short period of light exercise. A longer, more strenuous period of exercise, however, produced markedly different results; for two of the women S's (4 and 8) steadiness was increased, while for S 1 steadiness was decreased. For the three men S's, all showed a decrease in steadiness, S 19 being affected very little, S 17 to a greater extent, and S 2 to a much greater degree.

Again there is noted a peculiar situation when the performances of men and women are compared on Stimulus 4 (praise). The curve for S 17 (M) rose abruptly, S 19 went up slightly, and S 2 retained the same level as for Stimulus 3; S 8 (W) went down slightly following praise, S 4 rose slightly, and S 1 retained the same level as for Stimulus 3. The three curves for Stimulus 5 (reproof) for the men drop rapidly and to approximately the same extent. For the women the curve for S 1 rose considerably; S 4 fell slightly, and S 8 retained the same level as for praise. Sex differences again appeared in Stimulus 6. All of the curves for the three men rose; S 17 and S 19 went up but slightly, while S 2 rose to a marked extent. For the women, S's 1 and 4 registered a considerable increase in steadiness, while S 8 showed approximately the same decrease in steadiness.

After noting how the different stimuli affect the steadiness of the subjects one is interested in the relationship existing between the increase or decreases. In an attempt to secure an adequate measure of this factor, the correlation coefficients in Table II were computed. These were obtained by correlating the contact scores (Table I) for the different stimulus situations. The rank difference method of finding correlations was used and the resulting P's were transmuted into product-moment r's (5, p. 191). Only a few of these correlations may be considered substantial or high and by far the majority are low. It is observed that the few high correlations exist between the same type of stimulus situations; the correlations between exercise and praise and between exercise and reproof are low, indicating that changes in the emotional status of the subjects were due to the effect of the praise or reproof and not to the effect of exercise.

TABLE II

CORRELATION COEFFICIENTS BETWEEN STIMULUS SITUATIONS

Stimulus Situation	Men			Women		
	r	P.E.r	N	r	P.E.r	N
1-2	.79	.06	23	.85	.04	22
2-3	.48	.11	23	.76	.06	22
3-4	.24	.14	23	.31	.13	22
4-5	.37	.13	23	.60	.10	22
5-6	.39	.13	23	.42	.12	22
1-6	.42	.12	23	.84	.04	22
2-6	.25	.14	23	.77	.06	22
3-6	.50	.11	23	.86	.04	22
4-6	.29	.13	23	.28	.14	22
3-5	.30	.13	23	.66	.08	22
1-4	.39	.12	23	.32	.13	22
1-5	.76	.06	23	.46	.12	22

Group differences. A comparison of group performance offers a more satisfactory and perhaps a more reliable means of studying the changes in the emotional status of the subjects resulting from different stimulus situations. Studying the sex groups separately (Table I and Figure 3) it will be seen that the average number of contacts made for each stimulus situation by the men varied considerably from one stimulus situation to another. For purposes of group comparisons, it adds to the clearness of the discussion if the normal steadiness is: (1) compared with the changes in steadiness resulting from the exercise Stimuli 2, 3, and 6; (2) with praise and reproof; and (3) between exercise and praise and reproof.

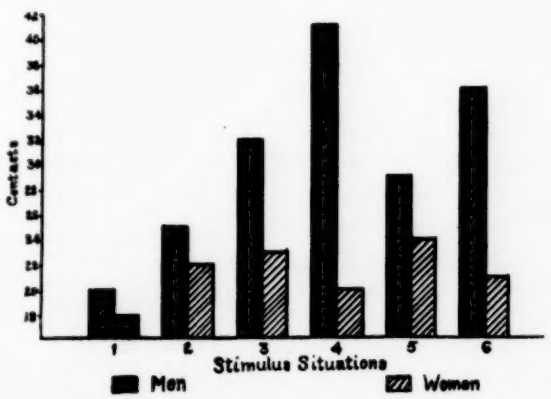


Figure 3. Average Number of Contacts Made by Men and Women S's for Each Stimulus Situation.

Men. (1) The average number of contacts for the norm stimulus was 20; for Stimulus 2, 25; for Stimulus 3, 32; and for Stimulus 6, 36. The difference in number of contacts between Stimuli 1 and 2 is 5; between Stimuli 2 and 3 is 7; between Stimuli 3 and 6 is 3; between Stimuli 1 and 6 is 16; and between Stimuli 2 and 6 is 11. The first point to be considered in relation to these differences is their reliabilities (5, pp. 128-133). In Table III the statistical reliability of the differences between the different stimuli is computed. Thus, it is seen that the differences between the norm and the three exercise stimuli are all statistically reliable. This shows that exercise (basketball) will produce a difference in steadiness, whether the period of exercise be of short, long, or medium duration.

(2) The average number of contacts made on Stimulus 4 was 41 and on Stimulus 5, 29. Comparing these averages with the norm (20) shows that steadiness was decreased by praise 21 contacts and by reproof 9 contacts. Referring to Table II, we find that these differences are reliable. A difference of 12 contacts was made between praise and reproof and this difference is also reliable. Both praise and reproof decrease the steadiness of the S's, and the decrease by praise and reproof is a reliable difference.

(3) The most interesting comparisons of differences between the various stimulus situations used are those between the exercise Stimuli 3 and 6, and Stimuli 4 and 5, reproof.[1] The difference between the average number of contacts for Stimuli 3 and 4 shows that steadiness which is decreased by nine contacts is reliable; between Stimuli 4 and 6 steadiness is increased, as is shown by a difference of five contacts. The obtained difference of five contacts is 2.00 times the standard deviation of the difference, 2.50, showing that the chances of praise (after a short period of exercise) causing more change in the emotional status of the learner than does a longer period of more strenuous exercise are 980 in 999+. There is a difference of three contacts between Stimuli 3 and 5 which is almost statistically reliable (997 in 999+), indicating that there is a tendency for steadiness to be decreased more by reproof than by a period of exercise of medium length. Between Stimuli 5 and 6 there is a difference of seven contacts--this shows there is virtually a true tendency for the long period of strenous exercise to cause a greater decrease in steadiness than does reproof.

Women. (1) The average number of contacts made by the women for each of the exercise stimuli is as follows: Stimulus 1, 18; Stimulus 2, 22; Stimulus 3, 23; Stimulus 6, 21. The differences between the averages of these stimuli are not so great as they were

1. The stimulus situations for 3, 4, and 5 were exactly the same except that praise was added to Stimulus 4 and reproof to Stimulus 5.

TABLE III

RELIABILITY OF THE DIFFERENCES BETWEEN DIFFERENT STIMULI

MEN

	Stimulus Situations											
	1	2	3	4	5	6	1	6	2	6	3	6
Ave.	20	25	32	41	29	36	20	36	25	36	32	36
S.D.	5.76	6.13	6.71	10.50	11.87	8.53	5.76	8.53	6.13	8.53	6.71	8.53
S.D. Ave.	1.20	1.28	1.40	2.19	2.47	1.78	1.20	1.78	1.28	1.78	1.40	1.78
Diff.	5	7	9	12	7		16		11		4	
S.D.diff.	.80	2.30	1.07	2.63	2.41		1.68		1.91		1.62	
D / S.D.diff.	6.25	3.91	6.54	4.56	2.90		9.52		5.76		2.47	
Chances in 1000 of real diff.	999+	999+	999+	999+	998		999+		999+		993	

	3	5	4	6	1	4	1	5
Ave.	32	29	41	36	20	41	20	29
S.D.	6.71	11.87	10.50	8.53	5.76	10.50	5.76	11.87
S.D. Ave.	1.40	2.47	2.19	1.78	1.20	2.19	1.20	2.47
Diff.	7		5		21		9	
S.D.diff.	2.45		2.50		2.05		1.74	
D / S.D.diff.	2.86		2.00		10.24		3.45	
Chances in 1000 of real diff.	997		980		999+		999+	

TABLE IV

RELIABILITY OF THE DIFFERENCES BETWEEN DIFFERENT STIMULI

WOMEN

	Stimulus Situations											
	1	2	3	4	5	6	1	6	2	6	3	6
Ave.	18	22	23	20	24	21	18	21	22	21	23	21
S.D.	6.33	9.68	6.34	5.45	6.50	7.92	6.33	7.92	9.68	7.92	6.34	7.92
S.D. Ave.	1.35	2.06	1.35	1.16	1.39	1.69	1.35	1.69	2.06	1.69	1.35	1.69
Diff.	4	1	3	4	3		3		1		2	
S.D.diff.	1.15	1.35	1.38	1.11	1.68		.85		1.31		1.76	
D / S.D.diff.	3.50	.74	2.17	3.60	1.79		3.53		.76		1.14	
Chances in 1000 of real diff.	999+	770	985	999+	960		999+		770		870	

	3	5	4	6	1	4	1	5
Ave.	23	24	20	21	18	20	18	24
S.D.	6.34	6.50	5.45	7.92	6.33	5.45	6.33	6.50
S.D. Ave.	1.36	1.39	1.16	1.69	1.35	1.16	1.35	1.39
Diff.	1		1		2		6	
S.D.diff.	1.13		1.76		1.40		1.42	
D / S.D.diff.	.88		.57		1.43		4.22	
Chances in 1000 of real diff.	820		7.15		925		999+	

TABLE V

GROUP VARIATIONS BETWEEN DIFFERENT STIMULI

	Increase in No. of Contacts				Decrease in No. of Contacts				Same	
	Men		Women		Men		Women		M	W
Stimuli	No.	Increase	No.	Increase	No.	Decrease	No.	Decrease	No.	No.
1 to 2	21	1 to 10	17	1 to 18	1	1	2	1 to 14	1	3
2 to 3	20	1 to 16	14	2 to 13	1	1	7	1 to 13	2	1
3 to 4	17	1 to 34	4	3 to 19	5	1 to 11	17	1 to 11	1	1
4 to 5	1	15	15	3 to 16	22	2 to 33	6	2 to 7	0	1
5 to 6	17	1 to 20	5	1 to 13	6	3 to 6	16	1 to 14	0	1
1 to 3	23	3 to 19	18	1 to 14	0	0	1	2	0	3
1 to 6	23	2 to 34	16	1 to 14	0	0	5	1 to 4	0	1
3 to 5	10	2 to 10	11	1 to 12	13	1 to 20	11	1 to 8	0	0
3 to 6	18	1 to 25	5	1 to 11	3	4 to 9	17	1 to 11	2	0
4 to 6	8	1 to 14	11	1 to 13	15	3 to 30	9	1 to 21	0	2

for the men (Table IV); however, the difference (four contacts) between the norm and Stimulus 2 (three contacts), and between the norm and Stimulus 6 are reliable. This shows that there is a true tendency for both the short and long periods of exercise in the water to decrease steadiness. The difference of one contact between Stimuli 2 and 3 shows that the probability of a longer period of exercise in swimming producing a decrease in steadiness is only slightly more reliable than chance. A much longer period of exercise in the water increased steadiness, but the obtained difference of two contacts is not sufficient to show a true tendency for this to occur.

(2) The average number of contacts made on Stimulus 4 was 20, and on Stimulus 5, 24. Comparing these averages with the norm (18) shows that steadiness was decreased but slightly by praise (two contacts), but to a more marked extent by reproof (six contacts). The obtained difference of two contacts between the norm and praise is not entirely reliable, but the reliability is high enough (925 in 999+) to show a tendency for praise to reduce the steadiness of the subjects; the difference of six contacts between the norm and reproof is entirely reliable, which fact shows a true tendency for reproof to reduce the steadiness of the subjects; the difference of four contacts between praise and reproof is statistically reliable, showing that there is a true tendency for reproof to decrease the steadiness of the S's to a greater extent than does praise.

(3) A comparison of the average number of contacts made on the exercise Stimuli 3 and 6 and Stimuli 4 and 5, praise and reproof, presents an interesting study. The difference between Stimuli 3 and 4 is three contacts, indicating that praise causes an increase in steadiness. The obtained difference of three is, however, only 2.17 times the standard deviation of the difference, 1.39, so the chances of this increase in steadiness representing a true tendency are 985 in 999+. The obtained difference of one contact between Stimuli 4 and 6 is sufficiently large to insure more than an even chance (715 in 999+) that a long period of strenuous exercise in the water will tend to reduce steadiness more than praise. The difference of one contact between Stimuli 3 and 5 is only .88 times the standard deviation of the difference, 1.13, so the chances of this increase representing a true tendency are 820 in 999+. Between Stimuli 5 and 6 there is a difference of three contacts, which indicates that long, strenuous exercise in the water will increase steadiness more than reproof. This increase represents a true tendency, 960 chances in 999+.

Another method of showing group variations in the effect of various stimulus situations upon steadiness is presented in Table V. This table shows the number of subjects whose steadiness was increased, remained

TABLE VI

RELIABILITY OF SEX DIFFERENCES

	Stimulus Situations											
	1		2		3		4		5		6	
	M	W	M	W	M	W	M	W	M	W	M	W
Ave.	20	18	25	22	32	23	41	20	29	24	36	21
S.D.	5.76	6.33	6.13	9.68	6.71	6.74	10.50	5.45	11.87	6.50	8.53	7.92
S.D. Ave.	1.20	1.35	1.27	2.06	1.40	1.35	2.19	1.16	2.47	1.39	1.78	1.69
Diff.	2		3		9		21		5		15	
S.D.diff.	1.80		2.42		1.94		2.48		2.83		2.46	
D / S.D.diff.	1.11		1.24		4.64		8.47		1.77		6.10	
Chances in 1000 of real diff.	860		885		999+		999+		960		999+	

the same, or was decreased by the various stimulus situations for which they were tested.

Sex differences. A comparison of the effect produced on the emotional status of the men and women groups shows: (1) that the same stimuli affected the steadiness of men and women to different extents and that (2) the women were steadier than the men. An examination of the steadiness curves (Figure 1) shows that for all six of the stimulus situations, the women were steadier than the men. The norm for the men was two contacts higher than for the women, which shows that the chances of this difference in steadiness representing a true tendency are 860 in 999+ (Table VI). A short period of light exercise decreases the steadiness of men more (three contacts) than of women. The chances that this difference is a true tendency are 885 in 999+. The obtained difference of nine contacts between the steadiness of men and women on Stimulus 3 is statistically reliable, showing that there is a true tendency for a period of more strenuous exercise to decrease the steadiness of men more than of women. Following praise the men registered a degree of unsteadiness of twenty-one contacts greater than the women, which is a true tendency for men to be less steady after praise than women. The steadiness of both the men and the women was affected by reproof, the men by five more contacts than the women. The chances that this difference rperesents a true tendency are 960 in 1000. The difference of fifteen contacts between the degree of steadiness for Stimulus 6 is entirely reliable.

An examination of Figure 3 shows that Stimuli 2 and 3 caused a decrease in steadiness of both men and women; that Stimulus 4 (praise) caused a marked decrease (nine contacts) in the steadiness of the men, while the steadiness of the women was increased three contacts. Exactly the opposite is shown for Stimulus 5 (reproof); the steadiness of the men was increased by three contacts while that of the women was decreased by four contacts. For Stimulus 6 the steadiness of the men was decreased, but not to the same extent as in the case of praise, while the steadiness of the women was increased as compared with the results in the case of reproof.

INTERPRETATION OF RESULTS

It is significant: (1) that the three different types of stimuli--exercise, praise and reproof--used in this investigation produced changes in the emotional status of the learner as is shown by the differences in steadiness; (2) that the same stimuli affected the emotional status of the subjects, producing different degrees of steadiness; (3) that the same stimuli produced reliable sex differences in steadiness; and (4) that the results possess a high degree of reliability.

The first three of these results compare favorably with the findings of other investigators who have studied the effects

of various stimuli on the emotional status of the learner. The last finding, however, is not substantiated by earlier investigators. Kellogg (11, p. 151) in discussing the effect of a "startle" type of stimulus on men and women says: "It is by no means apparent from our results that the women as a class were markedly more affected than the men." Skaggs (16) says that women are more emotional than men because they were more afraid of the electric shock than were the men. This conclusion, however, was based on introspective reports and upon the general observations of the investigator, and not upon a study of the objective data. A study of his tabulated figures indicates that the men were more upset than the women.

That women are more emotional than men seems to be the common opinion among laymen. The data here reported do not support this common belief and seemingly disprove the data of Skaggs.

It is to be remembered, however, that there were a number of factors which were not entirely controlled in this investigation, any one of which possibly could have been responsible for the sex differences. These partially uncontrolled factors were as follows:

1. The types of exercise participated in by the groups were of a different nature.
2. The length of the exercise periods differed:
   a. Two hours and thirty minutes for the men.
   b. Forty minutes for the women.
3. The praise and reproof were given by different persons:
   a. By the freshman basketball coach for the men.
   b. By the swimming instructor for the women.
4. There was a difference in the aims of the two groups:
   a. The men were striving to earn the freshman basketball numeral and the prestige that accompanies it.
   b. The girls were merely fulfilling a part of the physical education requirements of the university.

Such adjustments as were made in the attempt to control these factors were described in the explanation of the experimental conditions.

The individual changes in the emotional status of the men indicate that, in general, the subject with a high norm (less steady) also will rank high after a short period of light exercise, and that subject with a low norm (steady) will rank low after a short period of not too strenuous exercise. For a longer and more strenuous period of exercise, knowledge of the normal steadiness or lack of steadiness of the learner possesses little reliable predictive value (Table II). The data do not give the reasons for this condition. Whether the changes in emotional status are due to fatigue or to other physiological changes must be determined by other investigators.

Considering group performance, praise caused a decrease in steadiness in the men and reproof produced an increase in steadiness. The uniformity by which the subjects reacted to these stimuli and also the reliability of the obtained differences indicate that they possess valid predictive value.

The findings concerning the effect of praise and reproof on the steadiness of the subjects might be criticized on the basis that the men were generally praised on "off days" (when they were normally unsteady) or when for some other physiological reason they could not play as well as normally, thereby decreasing their steadiness, and that they were reproved on the days when they were at their best. The same criticism might be directed at the results secured from the women. Our testing procedure, however, would offset the probability that this criticism might be just and the correlations (Table II) between the different stimuli show that there was no preference of days for praise or reproof.

The data show that a short period of exercise in the water decreased the steadiness of the women, while a longer period of exercise tended to increase their steadiness to approximately a normal degree. The fact that a few minutes of exercise in the water decreased the steadiness of the women while a longer period of exercise increased their steadiness is not answered by the data. This

condition may be due to the shock (chill) of first entering the water, to adaptability, or to some other physiological factor. The findings on this point are, however, in harmony with the common opinion of swimming instructors and swimming participants who believe that swimming is the best type of exercise in which to participate because of its steadying effects on the "nerves" of the individual.

The sex groups as well as individuals of the group are affected to a different extent by praise and reproof. Women under the influence of a stimulus of praise are more steady than men. When reproved the conditions are exactly reversed; the women are less steady than men. The fact that reproof produced a steadying effect on the men might be due to the fact that the criticism was constructive, which acted more as a check on the learner's progress than it did to produce an effect on his emotional status. This, however, did not explain the differences existing in the different effects of praise and reproof on the sexes. The fact that women are steadier than men in every stimulus situation involved in this investigation seems to be an important finding, and the fact that these sex differences possess a high degree of reliability should add to the significance of the results.

The results of this investigation seem to point to many pertinent facts concerning basketball coaching. Whether the emotional status of the players is closely related to the length of practice periods, to the time a team should be on the floor before a game, when practice on "free throws" should be conducted, the length of scrimmage sessions, and other problems of a like nature, the author will attempt to determine by a more extensive investigation.

If subsequent investigations verify the results of this study, showing that swimming tends to produce an increase in the steadiness of the individual, then these results would be of considerable practical importance; swimming should be recommended for "nervous" individuals.

SUMMARY

To study the effect of various stimuli on the emotional status of the individual as expressed through changes in steadiness, 45 subjects-- 23 men participating in freshman basketball and 22 women participating in swimming--were given a series of steadiness tests, during which various degrees of exercise stimuli, praise stimuli, and reproof stimuli were presented without the learners being aware of the purpose of the investigation. An improved plate-and-stylus tester of the Whipple type was used to measure steadiness. Three types of exercise stimuli were presented: (1) a short period of light exercise; (2) a longer period of more strenuous exercise; and (3) a long period of strenuous exercise. The exercise for the men consisted of basketball practice, and for the women, swimming practice. The praise (4) and reproof (5) stimuli of a natural constructive nature were given under identical conditions.

1. The steadiness of the subjects was affected by each of the five stimulus situations.

2. The emotional status of men and women is affected to different degrees by the same stimuli.

3. Exercise in basketball affects the emotional status of men by decreasing their steadiness, whether the period of exercise is of long, medium, or short duration.

4. There is a wide difference in the normal steadiness of individuals.

5. Steadiness is decreased less by a long period of strenuous exercise in relation to the normal steadiness of the individual than by a medium period of exercise.

6. A short period of exercise in the water decreases the steadiness of women.

7. A longer period of exercise in the water produces little additional effect on the emotional status of women, but a long period of strenuous exercise in the water increases the steadiness of women to a degree closely approaching their normal steadiness.

8. The steadiness records, treated as a whole, possess an exceptionally high degree of reliability.

9. Both praise and reproof produce statistically reliable differences in the emotional status of men learners.

10. Steadiness is decreased less in men by reproof than by praise.

11. Praise affects the emotional status of men by reducing steadiness to a greater degree than does either a long or short period of exercise.

12. Reproof affects the emotional status of men by producing a higher degree of steadiness than that evidenced for either the medium or long exercise period.

13. Women are steadier than men under normal conditions.

14. Women are steadier than men under all stimulus situations used in this experiment.

15. Praise acts on the emotional status of women to increase their steadiness, while for men it decreases their steadiness.

16. Reproof increases the steadiness of men and decreases the steadiness of women.

17. An exercise period of medium length in basketball reduces the steadiness of men while a similar period of exercise in the water has little effect on the steadiness of women.

18. A period of long strenuous exercise in basketball decreases the steadiness of men while a similar period of exercise in the water tends to increase the steadiness of women.

19. Praise affects the emotional status of women by increasing their steadiness.

20. Reproof decreases the steadiness of women.

21. A period of short and medium length exercise in the water decreases steadiness of women, but a longer, more strenuous period tends to increase steadiness.

REFERENCES

1. Berry, Elmer. The Philosophy of Athletics (New York: A. S. Barnes and Company, 1927), pp. 147-150.
2. Bird, Charles. Effective Study Habits (New York: Century Company, 1931), p. 17.
3. Book, William F. Economy and Technique of Learning (Boston: D. C. Heath and Company, 1932), p. 100.
4. Chapman, J. C., and Feder, R. B. "The Effect of External Incentives on Improvement," Journal of Educational Psychology, VIII (1917), pp. ~~868-874~~ 413-414.
5. Garrett, Henry E. Statistics in Psychology and in Education, New York: Longmans, Green and Company, 1926.
6. Gates, Georgiana, and Rissland, Louise Q. "Effect of Encouragement Upon Performance," Journal of Educational Psychology, XIV (1923), pp. 21-25.
7. Gilchrist, E. P. "The Extent to Which Praise and Reproof Affect a Pupil's Work," School and Society, IV (1916), pp. 872-874.
8. Griffith, Coleman R. Psychology of Coaching (New York: Scribner's Sons, 1929), p. 198.
9. Hurlock, E. B. "An Evaluation of Certain Incentives Used in School Work," Journal of Educational Psychology, XVI (1924), pp. 145-149,
10. Hurlock, E. B. "The Value of Praise and Reproof as Incentives for Children," Archives of Psychology, No. 71 (1924), p. 78.
11. Kellogg, W. N. "The Effect of Emotional Excitement Upon Muscular Steadiness," Journal of Experimental Psychology, XIV (1932), pp. 142-166.
12. Kirby, T. F. Practice in the Case of School Children. Teachers College Contributions to Education, No. 58. New York: Teachers College, Columbia University, 1913.
13. Kitson, H. D. "A Study of the Output of Workers Under A Particular Wage Incentive," University Journal of Business, Vol. I, pp. 54-68.
14. Landis, C., and De Wick, H. N. "The Electric Phenomena of the Skin (Psycho-Galvanic Reflex)", Psychological Bulletin, XXVI (1929), pp. 64-119.
15. Sims, V. M. "The Relative Influence of Two Types of Motivation on Improvement," Journal of Educational Psychology, XIX (1928), pp. 480-484.
16. Skaggs, E. B. "Changes in Pulse, Breathing, and Steadiness Under Conditions of Startledness and Excited Expectancy," Journal of Comparative Psychology, VI (1926), pp. 303-317.
17. Symonds, Percival M. Diagnosing Personality and Conduct. (New York: Century Company, 1931), p. 377.
18. Thorndike, Edward L. Human Learning (New York: Century Company, 1931), chapters 3, 4, and 5.
19. Whipple, G. M. Manual of Mental and Physical Tests (Baltimore: Warwick and York, 1924), p. 157

MEASURING ATTITUDE TOWARD THE HIGH SCHOOL [1]

by
H. H. Remmers,
G. C. Brandenburg,
and
F. H. Gillespie
Purdue University

INTRODUCTION

The proposition that desired educational outcomes must somehow be measured rather than assumed no longer needs defense among competent educators. The fallacy inherent in the converse proposition has been too abundantly demonstrated in the experimental literature to require any extended argument here. Whether the scientific method guided by an adequate social-educational philosophy be a sufficiently trustworthy tool, is still a matter of some debate among sincere and thoughtful educators, but any alternative is, we believe, in the nature of a retrogression and a council of despair.

Among the significant outcomes of educational procedures pupil attitudes have received a great deal of attention by way of discussion in the educational literature, but attempts at measurement have been notably scanty. An enormous body of literature has accumulated within the past two decades or so around the problem of measuring achievement in the traditional subject-matter of the curriculum, so that in this sector of educational Wissenschaft most rapid and gratifying progress has been made. In the case of presumably less "tangible" outcomes or objectives such as attitudes, however, the situation, while distinctly promising, is far less satisfactory.

The reasons for this state of affairs are somewhat complicated and obscure. There are, or at least have been, the sheer technical difficulties of the theoretical development, definition, and practical construction of adequate measuring instruments for phenomena as complex as those subsumed under the term "attitudes." The rapid development of statistical tools and especially the more recent work in psychology related to this problem have made possible rapid advances. Besides these technical difficulties, however, the problem is beset by the inertia, not to say recalcitrance, which exists with reference to the invasion of what in psychology have been called the emotions. The measurement of affectivity is made difficult by the existence of a variety of idols of the den and the market place, to use Bacon's picturesque but highly descriptive terminology. "Vested interests"--sentimental, social, economic, traditional--in the form of folkways have stood in the way of such measurement, and will likely continue to act as a brake upon progress in this direction, particularly with reference to educational, social, and economic changes which may be indicated as the result of such measurement.

But even within a very narrowly defined status quo representing the resultant of present educational and social lines of force, so to say, there is still possible a considerable measure of science applied to the measurement of affectivity. That is, there are attitudes presumably having their origin in the social-educational continuum, upon which all individuals concerned would agree as being desirable of measurement. The attitude of the individual student toward the high school as such we conceive to be such an attitude. In the nature of the

1. Division of labor in this study is as follows. The problem was suggested by Remmers who is also responsible for the writing of the present paper. Gillespie under the joint direction of Brandenburg and Remmers constructed the attitude scale described in this paper as a master's thesis. The reliability of the scale was computed by Remmers from data supplied by Gillespie and Delia L. Silance.

case it will be safest to begin a program of measurement which shall include attitude as a significant variable with one free from possibly unpleasant repercussions. The scale to be described below is the first of a series that is planned for experimental use, particularly in educational situations.

The attitude scale is based upon the psycho-physical principle that equally often noticed differences are equal, a principle very fruitfully applied to the scaling and measuring of attitudes by L. L. Thurstone and his students. In the references given at the end of this paper the interested reader who is not familiar with the details of the underlying theory and procedures will find adequate discussion of these. Such details are omitted from this paper in the interests of space conservation. Suffice it to say that affectively toned debatable statements which an individual endorses are assumed to constitute a valid index of the individual's attitude toward the object defined by the attitude scale, in the present case attitude toward the high school. Following are the two forms of the scale as they were used in investigations reserved for later publication elsewhere.

The scale values (medians) and Q values (interquartile range) for the statements which were finally included in the scale on the basis of experimental evaluation with 100 subjects are given in Table I. These 100 subjects sorted 110 affected statements. From these, 50 were retained for two forms of the scale, on the basis of two criteria: (1) scale values covering approximately the entire scale (eleven points) and (2) small Q values, showing absence of ambiguity.[1]

The reliability of the scale was determined from two independent samplings in two different high schools. One sampling of 179 pupils, approximately two thirds of the pupil population in a small town high school, yielded a reliability coefficient for one form of .753 ± .02. Another sampling of 300 from a total high school population of approximately 1,250 in a city of about 30,000 inhabitants gave a reliability of .727 ± .02. These reliability values become .859 ± .02 and .841 ± .02 respectively when stepped up by the Spearman-Brown formula for the reliability of the sum of the scores on both forms of the scale. This value compares favorably with the reliability of most standardized educational measuring instruments.[2]

The scale is scored by averaging for each subject the scale values of all items which he had placed a check mark, a procedure quickly accomplished by the aid of an adding machine and a stencil on which appear the scale values corresponding to each of the items in the scale.

TABLE I
SCALE VALUES (MEDIAN) AND VARIABILITY VALUES (INTERQUARTILE RANGE) DETERMINED FOR TWO FORMS OF THE SCALE FOR MEASURING ATTITUDE TOWARD THE HIGH SCHOOL

Statement Number	Scale Value		Interquartile Range Value	
	Form A	Form B	Form A	Form B
1	7.1	0.6	2.2	1.8
2	8.8	6.4	1.6	1.2
3	3.2	7.2	2.5	2.4
4	3.5	8.1	2.4	1.5
5	9.0	7.3	1.8	1.8
6	2.4	4.5	1.7	2.5
7	2.2	2.2	2.3	2.4
8	5.3	6.1	1.6	1.6
9	6.8	9.2	1.7	1.8
10	7.2	5.9	1.3	1.1
11	9.4	4.4	2.2	2.2
12	7.2	8.8	2.8	1.7
13	6.0	8.9	0.9	2.2
14	4.5	3.6	2.9	2.2
15	2.8	2.7	2.5	1.9
16	1.4	2.9	2.4	2.0
17	0.6	5.8	1.8	1.9
18	1.1	5.6	1.8	0.4
19	4.4	1.3	1.9	2.6
20	10.8	3.8	1.8	2.1
21	5.7	2.0	1.0	1.9
22	3.5	10.1	1.9	2.0
23	1.5	1.2	2.0	2.2
24	9.8	9.7	2.0	1.6
25	9.5	9.4	2.0	1.3

1. Gillespie's thesis gives the details of the experimental procedure in scale construction and contains the scale and Q values for all the original 110 statements. The thesis is on file in the Purdue University Library.
2. In another place is discussed the difference in reliability coefficients to be expected from two different methods of determining reliability—the "split-test" and "equivalent forms." See H. H. Remmers, "A Possible Experimental Error in Determining the Overlap of Two Correlated Variable," to appear in an early number of the Journal of General Psychology. The theoretical argument made in this article has since been verified experimentally, and will be published shortly. The reliabilities reported for many analogous educational measuring instruments have been determined by the split-test procedure and are without doubt in general spuriously high.

HIGH SCHOOL ATTITUDE SCALE

Form A

Please fill in the blanks below. (You may leave the space for your name blank if you wish.)

1. Name .....
2. Age when school started this year .....
3. Grade in school .....
4. Boy or girl .....
5. Occupation of parent or guardian .....
6. What life work do you expect to enter? .....
7. Have you ever failed in a school subject in the grades? ..... In high school? ..... If so, in what subjects? .....

Below is a list of twenty-five statements about school. Place a check mark before each statement with which you agree, and leave unmarked those with which you disagree. This test will in no way affect your standing in school.

..... 1. I would rather go to school than work.
..... 2. I like to do school work.
..... 3. My classes are very uninteresting.
..... 4. It costs too much to go to school.
..... 5. It helps one to get a job if he has a high school diploma.
..... 6. I can learn more working on a job than in high school.
..... 7. A high school graduate is often worse off morally than he was before going to high school.
..... 8. High school is all right, but I don't like it.
..... 9. High school has its drawbacks, but I like to go.
.....10. I would rather go to high school than to stay at home.
.....11. A high school education will help one to be a good citizen.
.....12. I don't like to associate with people who haven't a high school education.
.....13. I don't care about high school, but I think one ought to.
.....14. High school may be all right, but I don't think it does any good.
.....15. Most people would be just as well off not to go to high school.
.....16. High schools are a waste of the taxpayers' money.
.....17. I think every high school in the country ought to be closed and the teachers put to work at something useful.
.....18. High school teachers are parasites on the community.
.....19. If one's parents need him at home he ought not to go to high school.
.....20. A high school education is worth a million dollars to any young person.
.....21. Sometimes I like high school and sometimes I don't.
.....22. There are too many rules and regulations in schools.
.....23. In high school pupils learn to disrespect everything of high idealistic character.
.....24. The high schools do more to develop high standards of conduct than any other institution.
.....25. High school develops self reliance.

Place a check (√) below along the line to indicate your attitude towards school. A is extremely favorable and K is extremely against. F is neutral and the other letters indicate graded steps between.

A B C D E F G H I J K

HIGH SCHOOL ATTITUDE SCALE

Form B

Please fill in the blanks below. (You may leave the space for your name blank if you wish.)

1. Name .....
2. Age when school started this year.....
3. Grade in school.....
4. Boy or girl.....
5. Occupation of parent or guardian.....
6. What life work do you expect to enter?.....
7. Have you ever failed in a school subject in the grades?.....
   In high school?..... If so, in what subjects?.....
   .....

Below is a list of twenty-five statements about school. Place a check mark before each statement with which you agree, and leave unmarked those with which you disagree. This test will in no way affect your standing in school.

..... 1. School is like a prison.
..... 2. I have a lot of fun in school.
..... 3. School is all right.
..... 4. My teachers always treat me fairly.
..... 5. I like to go to school to be with other people.
..... 6. Many of our great men have no high school education.
..... 7. I hate most school work.
..... 8. Some things about high school are all right.
..... 9. High school training develops personality.
.....10. High school is a good thing for some people and a bad thing for others
.....11. I would just as soon stay at home as go to school.
.....12. All the better class of people have high school educations.
.....13. The high schools lift the plane of sportsmanship in a community.
.....14. Too much money is being spent on the high schools for the benefit received.
.....15. The high school teaches mostly old useless information.
.....16. They won't teach things one really wants to know in high school.
.....17. If one has plenty of money it may be all right to go to high school.
.....18. I haven't any definite like or dislike for high school.
.....19. Any old fogy knows more than a high school graduate.
.....20. The kindest and best people I know don't have high school educations.
.....21. High school cramps and dwarfs one's personality.
.....22. America could not stand as a nation if it were not for our high schools.
.....23. Our high schools teach immorality and indecency.
.....24. High school training develops high ideals in the pupils.
.....25. High school develops loyalty.

Place a check (√) below along the line to indicate your attitude towards school. A is extremely favorable and K is extremely against. F is neutral and the other letters indicate graded steps between.

A B C D E F G H I J K

BIBLIOGRAPHY

1. Bain, R. "Theory and Measurement of Attitudes and Opinion," Psychological Bulletin, XXVII (May, 1930), pp. 357-80.
2. Chave, E. J. "A New Type of Scale for Measuring Attitude," Religious Education, XXIII (1928), pp. 364-69.
3. Hartshorne, H., and May, M. "First Steps Toward a scale for Measuring Attitude," Journal of Educational Psychology, XVII (1926), pp. 145-62.
4. Hinckley, E. D. "Influence of Individual Opinion on Construction of an Attitude Scale," Journal of Social Psychology, III (August, 1932), pp. 283-96.
5. Murphy, G., and Murphy, Lois B. Experimental Social Psychology. New York: Harper and Brothers, 1931. Especially Chapter XI, pp. 615-94, on "Social Attitudes and Their Measurement."
6. Peterson, Ruth C., and Thurstone, L. L. The Effect of Motion Pictures on the Social Attitudes of High School Children. Ann Arbor: Edwards Brothers, 1932. Pp. 1-69.
7. Silance, Ella Belle. The Construction of a Scale for Measuring Attitude Toward High School Subject X. Unpublished Master's Thesis, Purdue University, June, 1933.
8. Thurstone, L. L. "Theory of Attitude Measurement," Psychological Review, XXXVI (1929), pp. 222-41.
9. Thurstone, L. L. "Attitudes Can Be Measured," American Journal of Sociology, XXXIII (1928), 529-54.
10. Thurstone, L. L., and Chave, E. J. The Measurement of Attitude Toward the Church. Chicago: University of Chicago Press, 1929.
11. Thurstone, L. L. "A Scale for Measuring Attitude Toward Movies," Journal of Educational Research, XXII (September, 1930), pp. 89-94.
12. Wong, C.K. "Suggested Criteria for Writing Attitude Statements," Journal of Social Psychology, III (August, 1932), pp. 367-73.

THE MEASUREMENT OF ATTITUDES TOWARD CURRENT POLITICAL AND ECONOMIC PROBLEMS AMONG JUNIOR AND SENIOR HIGH SCHOOL PUPILS

by
Milo L. Whittaker[1]
Northern Illinois State Teachers College

PROBLEM FOR INVESTIGATION

The purpose of this study is to discover what attitudes school children possess in regard to current political and economic problems; to determine the effect upon attitude of different environmental factors such as a rural, town, or city background; to discover the relation of attitude to socio-economic status and to general intelligence; to investigate the relation of attitude to school grade; and finally to suggest what consequences for civic training might be deduced from discoveries resulting from this study.

Although there is some disagreement as to the meaning of the term "attitude", in this study it will be used to denote "the sum-total of man's inclinations and feelings, prejudice or bias, preconceived notions, ideas, fears, threats, and convictions about any specific topic" (14, p. 6-7).

Students of social psychology are not in complete agreement as to the practicability of measuring attitude. Some take the position that verbalized responses do not in any real sense reveal attitude. The contention of Bain (2) is that people tend to respond in conventional ways even though they do not have to sign their names, that most thinking is rationalization or stereotyped reaction, that questions are interpreted differently, and that words often refute action and vice versa. The fact that a high correlation frequently exists between successive answers to the same question does not establish the reliability of the method. On the other hand, Katz and Allport (5, p. vii-xi) believe that if the kind of information sought is such that a frank answer will not embarrass the subject, if the questions are so worded that the meaning is clear, and if the results are interpreted with the full understanding that responses to certain types of questions may lack candor, the attitude questionnaire may be considered a legitimate device in research.

Perhaps the clearest statement as to the possibility of measuring attitude has been presented by Thurstone and Chave (14). These authors recognize at the outset the complexity of an attitude and the fact that it cannot be completely described by any single numerical index. This may be said, however, of any measuring device, whether for measuring a table or the temperature of a room. "But this discrepancy between the index and 'truth' is universal" (14, p. 8).

Although a subject may be hiding the truth and practicing deliberate deceptions, it is nevertheless desirable to know what people claim to believe about certain questions. It is pointed out by both Katz and Allport and by Thurstone and Chave that we should not assume that an overt act is conclusive proof that we have discovered a person's true attitude. A moment's consideration should convince us that people often deceive by action as well as by words. The overt act of the politician in the presence of his constituency, the sanctimonious religionist praying before the altar, or the salesman in the presence of a prospective customer are probably no more accurate revelations of attitude than responses to a questionnaire.

> "All that we can do with an attitude scale is to measure the attitude actually expressed with the full realization that the subject may be consciously hiding his true attitude or that the social pressure of the situation has made him really believe what he expresses. This is a matter for interpretation. It is probably worth while to measure an attitude expressed by opinions. It is another

1. This study was directed at the Iowa Child Welfare Research Station by Dr. George D. Stoddard.

> problem to interpret in each case the extent to which the subjects have expressed what they really believe. All that we can do is to minimize as far as possible the conditions that prevent our subjects from telling the truth, or else to adjust our interpretations accordingly" (14, p. 10).

In brief, the possibility of measuring attitude depends upon the ability of the investigator to present a situation which will in no sense embarrass the subject, to place before him questions about which he may be expected to have an opinion and which are stated in such language as may be clearly understood.

Investigators should not become unduly disturbed by the discovery that overt behavior does not always correspond to the results of attitude tests. Allport's (1) study of 375 Dartmouth students shows that about one half of the students who were found to be political radicals by the test failed to vote for the radical presidential candidate. Many of us need not go beyond our own experience to be able to assert that radical tendencies in politics do not always find expression at the polls. The pronounced radical whose political ideas have crystallized will vote the radical ticket, but there are no doubt many others who for various reasons do not express their radicalism at the polls. Nevertheless they may be classed as possessing radical political attitudes. It is a tribute to Allport's test that it discovered twice as many radicals as were found at the polls. On the other hand, Porter (10) found that students who had overtly identified themselves with pacifist activities scored near the pacifist end of the scale, while R.O.T.C. officers scored near the opposite end.

Neumann (7) applied his attitude test on international relations to a number of army officers, members of the Stock Exchange on Wall Street, a number of declared pacifists, and other groups known to be pronounced in their international attitudes. The results were highly consistent. In view of the fact that it is a matter of common observation that one's agreement or disagreement with opinions of a controversial nature affect his behavior, "there seems to be no reason why this behavior should suddenly become non-significant when it is made the subject of careful inquiry." (6)

Most of the studies of attitude toward political and economic questions which have hitherto been made deal with college students or other adults. Little has been done toward investigation of the age when attitudes toward such problems begin to take form. Whether political and economic attitudes change significantly as children pass from grade to grade, and if so, in what direction, are problems worthy of careful investigation. Are these attitudes affected to a measurable extent by social backgrounds? Is there a "rural mind" and an "urban mind" which may be detected by attitude tests? To what extent does the literature of the social studies to which children are exposed in school affect their political and economic attitudes? These are the chief problems which constitute the present investigation.

METHOD OF STUDY

Field to be Investigated

This investigation was conducted along three lines, namely, attitude toward the present form of government in the United States, attitude toward capitalism as expressed in our present economic system, and attitude toward the labor group. The study is concerned primarily with three distinct groups of pupils in grades seven to twelve inclusive, pupils living in a city of 20,000 population, those living in small towns from 300 to 600 population, and the third group consisting of pupils living in the open country.

Care was taken to include in the study only children who had lived at least three years in the group in which they were studied. As a matter of fact, however, few cases were included who had not spent all or most of their lives in the group in which they were classified. A total of 612 pupils was studied, with slightly more than 200 in each group and at least 25 in each grade.

One hundred seventy-six pupils in grades seven, eight, and nine, who have had from four months to three year's study of the Rugg social science books, were used in a subsidiary survey. This survey was made

for the purpose of discovering what difference, if any, might be found in attitudes of this group as compared with other pupils.

The tests used in this study were administered during the month of December, 1932.

Constructing the Test[1]

The first step in the investigation consisted of the construction of an attitude test. The Watson (15) technique of measuring fairmindedness, with certain modifications, was used. Watson's test is in five parts, three of which have been used in this study. These three parts are designated as Tests A, B, and C, respectively.

Test A, which is called a generalization test, consists of twenty-four statements such as, "Members of labor unions are dangerous persons," "Laws are more favorable to the rich than to the poor," and "Millionaires came by their wealth honestly." Each of the twenty-four statements is preceded by the words All, Most, Many, Few, No. The pupil is instructed to underline the word which makes the statement one which he is willing to accept. Instead of being asked to mark the statement either true or false, he has the choice of five different responses. The purpose of this test is to measure the pupil's tendency to make broad generalizations or overstatements.

Test B, which is called degree of truth test, is a measure of intensity of opinion. Here the pupil is confronted by twenty-one statements similar to those in Test A, but each statement is preceded by the numerals 1, 2, 3, 4, and 5. The pupil is instructed to underline 1 if he feels that the statement is completely or certainly true; underline 2, if probably true; underline 3, if not certain or if it is an open question. If probably false, underline 4; and underline 5 if the statement is believed to be completely or certainly false.

Test C, which is called testing the meaning of facts, is designed to reveal the tendency of pupils to draw unwarranted conclusions from facts. This test consists of a series of eight statements which the pupil is told should be accepted as true and correct. Each statement is followed by a group of four conclusions or opinions introduced by the statement, "The above statement shows." The pupil is to mark with a plus sign each of the four conclusions or opinions which he is willing to accept. He is to use a minus sign for each one which he is unwilling to accept, and a zero for each one about which he is unable to make a decision. An illustrative statement together with the opinions or conclusions is presented below:

Statement of Fact

> **The average net income of all corporations in the United States since 1921 was approximately eight billion dollars. These incomes averaged less than five billion dollars from 1913 to 1915.**

The above statement shows:

> **1. That the corporations are whole-heartedly serving the public interests under the direction of the best and most intelligent citizens.**
>
> **2. That the corporations are robbing the people.**
>
> **3. That large corporations force their employees to work harder and for less pay than in small industries, thus increasing their profits.**
>
> **4. The economic development of the country during the past few years has greatly favored production by corporations to the disadvantage of the small producer.**

The technique employed in Test C differs materially from that of Watson in that his test presents an opportunity for marking but one of the several possible reactions to each statement of fact, and that marking must be an acceptance. In the present test the pupil is given an opportunity to react to each conclusion or opinion either by accepting, rejecting, or admitting that he cannot decide. A further difference lies in the fact that in the Watson test there is at least one neutral statement under each statement of fact. This neutral statement is merely a colorless reiteration of the statement of fact. Since the pupil is to mark but one of the conclusions or opinions under each statement of fact, he may easily

1. See appendix for the directions used in administering the test. Due to limitations of space the test itself is not reproduced.

take refuge behind the colorless statement, thereby concealing his attitude toward the problem. Instead of measuring the pupil's attitude of fairmindedness, it is possible that it measures only his ability to evade the issue. Test C, constructed by the writer, avoids this situation by requiring the pupil to react to each item.

It is important to note that none of the tests used in this study force the pupil into a position of expressing an opinion either stronger or weaker than is represented by his actual convictions. With a five-point gradation for his various reactions in Tests A and B, and with an opportunity to accept, reject, or register inability to decide in Test C, he is quite thoroughly safeguarded.

Selecting and Classifying the Content Material of the Tests

The statements comprising the three tests are based on the Rugg (11, 12, 13) social science textbooks. Because of the unusually thorough research and experimentation upon which the Rugg series is based, these books seemed to be a very desirable source from which to obtain the material for the tests. A rather complete list of the topics covered by these books was made and from these topics the statements comprising the preliminary tests were prepared. The preliminary tests contained about twice as much material as was desired for the final tests. The surplus material was eliminated in a manner to be explained later.

In order that the tests should measure attitude in the three fields previously named, attitude toward the government, toward the present capitalistic system, and toward the labor group, it was necessary to have all items in the tests classified under these three heads. The items in Test A and Test B were classified by fifty-five graduate students in sociology at the State University of Iowa, while the material in Test C was classified by forty-seven college teachers. All items were eliminated from the test upon which less than 60 per cent of the judges agreed as to classification.

Evaluating the Test Items

In order that the results of an attitude test may be expressed in quantitative terms with a degree of accuracy which is at all satisfactory, it is necessary to assign a numerical value to each possible reaction presented in the test. We may say that one object possesses more beauty than another object, or that this person is more radical than that person. Such statements are in a crude sense measurements of attitude, but the results of such measurements are not to be regarded as possessing much practical value. What we desire to know is: How much more beauty does this object possess than another? How much more radical is this person than that? When these "more" or "less" concepts are thought of as being represented on a linear scale, it is possible to assign numerical values to them.

The items in Test A and Test B were evaluated by a modification of the Thurstone technique (14). Each statement in Test A of the preliminary test was mimeographed upon a separate ticket, once using the "all" prefix and again using the "no" prefix. These tickets were given to groups of students in both graduate and undergraduate classes, who by following carefully prepared instructions[1] were to place each statement somewhere upon an eleven-point scale. This placement would represent the degree of favorableness or unfavorableness toward the field represented by the statement. The center of the scale, No. 6, indicated neutral or nearly neutral attitudes; the extreme right, No. 11, most favorable attitudes; and the extreme left, No. 1, most unfavorable attitudes. The judges placed all other tickets on the scale in accordance with the degree of favorableness or unfavorableness indicated by the statement. It should be noted that the judges were not asked to indorse or reject any of the statements contained on the tickets. Their task was simply to evaluate the statements made. The tickets were then banded and numbered by each judge so that in tabulating them each ticket would be given the place assigned by the judge.

1. See Appendix.

The scale value of each statement was determined graphically by means of a curve of accumulative proportions. The X-axis indicates the direction and degree of attitude and the Y-axis represents the accumulative proportions. The point at which the curve crosses the 50 per cent line indicates the scale value of the statement based on the median opinion of the judges. The measure of ambiguity of the statement is indicated by the distance between the two quartile points of the curve, i.e., the distance on the scale between the scale value at which the curve has an ordinate of .25 and its scale value at which it has an ordinate of .75. This distance is called the Q-value.

A statement with a large Q-value indicates that the judges did not agree well upon the scale value of the statement. If the Q-value is too large, the statement possesses little value as a measurement of attitude. Just how large the Q-value should be before a statement should be eliminated from the test is a matter of opinion. Thurstone (14) kept no statement whose Q-value was greater than 3.0. The Q-values in Thurstone's experiment were kept low, in part at least, by discarding the work of certain judges if there was strong indication of carelessness. By this means the work of slightly more than 12 per cent of his judges was discarded from his tabulated results.

Although there was considerable evidence of confusion in the minds of the judges who evaluated the material comprising the tests under consideration, nevertheless, no reports were discarded. In consequence of this, the Q-values are on the whole larger than in Thurstone's experiment. The average Q-value was found to be 3.38.

For purposes of scoring, the scale values referred to above were translated into scores by determining the distance in steps from the neutral point on the scale, namely 5.5, to the point represented by the scale value. Thus, if a scale value were 8.6 on the eleven-point scale, it would indicate a score of 8.6 minus 5.5 or a score of 3.1 in favorableness toward the field under investigation. A scale value of 1.2 would indicate a score value of 5.5 minus 1.2 or a score of 4.3 in unfavorableness. All scores are stated as either plus or minus, depending upon their direction from the neutral point.

After having obtained score values for the two extreme positions in each statement, it was necessary to assign values to the three middle statements. Instead of determining their values in the manner described above, a process which would have so overburdened the judges as to make the procedure impractical, the assumption was made that the scores of these positions in the test would appear as a simple mathematical distribution between the two extreme positions. To illustrate, Statement No. 2 in Test A is "Criminals in our large cities are shielded by officials." The "All" statement was given a score value by the judges of -4.3 while the "No" statement was given a score of 4.5. The distance between these two extreme positions on this statement is 8.8. Since there are four steps from "All" to "No", the score value to be assigned to "Few" will be diminished by 2.2 points (i.e., $\frac{8.8}{4}$), or 2.3. The value assigned to "Many" will be 2.3 - 2.2. or 0.1. The statement "Most" receives its score in the same way.

Score values for Test B were determined in the same manner except that the mimeographed tickets placed in the hands of the judges contained the direct statements as given in the test and an equal number of tickets in which the opposite of this statement was expressed. Thus, the first statement in Test B is, "The laws are more favorable to the rich than to the poor," while the opposite statement was, "It is not true that the laws are more favorable to the rich than to the poor."

Tests A and B were evaluated by 116 judges. After an evaluation by fifty-five judges had been made, the scale values were computed. They were computed again after the number of judges had been increased to 116. This seemed a sufficient number since the average scale value was changed only 0.1 points by the addition of the last sixty-one judges.

The nature of Test C is such that its items could not well be evaluated by the Thurstone technique. The score values of the various items in this test were assigned upon the basis of the judgment of forty-seven teachers, forty-four of whom were college

teachers and three were teachers in high school. The procedure was as follows: One set of twenty-five copies of Test C was marked by the writer as might have been marked by a pupil except that none was marked with a zero, which indicates inability to make a decision. Another set of twenty-five was marked just the opposite of the first set. These fifty copies were then sent to the fifty judges, one copy to each judge, together with a copy of mimeographed instructions for evaluating the test items.

The judges were to determine, first, the field to which each item in the test related, whether Government, Capital, or Labor, and next, to determine whether the item indicated an attitude of favorableness or unfavorableness toward this field. Complete returns were received from forty-seven judges. All items were eliminated from this test upon which fewer than 60 per cent of the judges agreed. Score values were weighted on the basis of the percentage of agreement of the judges as follows:

Per Cent Agreement	Score Value
60--69	1.0
70- 79	1.3
80- 89	1.6
90-100	2.0

Validity and Reliability

Since there are no standardized tests in this particular field, the validity of the test could not be determined by correlating it with other tests. Its validity rests chiefly upon the source from which the content of the test was taken, namely, the Rugg social science textbooks. As is generally known, the Rugg texts are based upon what is probably the most extensive research that has yet been conducted on what should constitute the content of the social studies.

The content of the Rugg texts was analyzed by the writer and a list of more than 100 problems was made. The present writer discarded problems not germane to the immediate interest in Capital, Labor, or Government. The list was further reduced by a committee of the graduate faculty, under whose direction the study was begun. The procedure of sorting the tickets by the judges who evaluated the material in the tests constituted a further validation. The judges were instructed to discard all statements which in their opinion did not pertain to the subject under investigation. The procedure employed in determining the numerical values of the various test items, which resulted in discarding much of the original material, constituted the last step in determining the validity of the test.

The reliability of each test was determined by the self-correlation method. The odd-numbered items were correlated with the even-numbered and the Spearman-Brown formula was applied to determine the reliability. The reliabilities were found to be as follows: Test A, .66 $\pm$ .04; Test B, .64 $\pm$ .04; and Test C, .61 $\pm$ .04. These reliabilities restrict the application of the tests to group comparisons.

PRESENTATION AND DISCUSSION OF MEAN SCORES

The two prime tasks in this section of the report are to present the results of the application of the tests, which have been described under the heading of method, on the basis of mean scores obtained and to discover what significant relationships, if any, exist among the scores of the various groups.

A study of these mean scores may be made in two different ways. The mean plus scores of the various groups may be computed and compared with one another to discover significant trends among those groups which show a favorable attitude toward the field under investigation. Then, the means of the minus scores may be treated in the same manner. This would give at least a partially satisfactory picture of the groups under consideration. The other alternative is to present the findings in terms of the means of net scores, i.e., the algebraic sum of the plus and minus scores. Since our chief concern is with total group attitude, it seemed desirable to adopt the second alternative and to deal chiefly with net scores.

The reader should recall that the investigation relates to three separate fields toward which attitude is being measured, namely, the present form of government, the

present economic order, and the labor group. A plus score is always interpreted as expressing a favorable attitude and a minus score an unfavorable attitude toward the field under investigation. For convenience of expression, the three fields will be referred to as Government, Capital, and Labor respectively.

By way of introducing the findings it may be said that each of the three groups, city, town, and rural, showed a pronounced attitude favorable to the Government in Test A as well as in Test B. (Owing to the fact that so many items in the preliminary material for Test C were eliminated through failure of the judges to agree, it was found that there was insufficient material in Government and Labor to constitute a satisfactory test. Hence only the scores in Capital are considered in Test C.) Each group also showed a favorable attitude toward Labor in both tests.

In Capital, however, there was a very close division of scores, especially in Tests A and B. In Test C the mean is in favor of Capital but by a relatively small margin. Because of this close division, it is not surprising that one group may be found with a mean score in favor of Capital and another with a mean score in the opposite direction, but both hovering around the zero point.

The study indicates few significant differences in attitude among the three geographical groups, except in Capital. Where these differences are large enough to appear at all significant, they will be discussed in this chapter. The complete data for this part of the study may be found in Tables I and II.

The chief fact of importance in connection with all tests is the uniformity of mean scores from group to group. The rural group, however, stands out in certain of the comparisons in such a way as to give some basis for asserting that, as a group, the rural pupils exhibit attitudes which separate them from the other groups. The difference which this group exhibits is toward a stronger opposition to our economic system. The city group stands out somewhat in its support of our economic system, being the only one to obtain a mean score favorable to Capital.

Attention should be called to the fact that the mean scores in Test C are all favorable to Capital. It seems probable that this is due to the nature of the test which does not allow for the more careful choosing of positions on the various statements, and also because the method of scoring does not permit those finer numerical discriminations which are characteristic of Tests A and B. The rural group shows a more unfavorable attitude toward Capital in all tests than either of the other groups. These differences are great enough to be significant.

The most important statement that may be made concerning the town group thus far is that it exhibits a tendency to stand midway between the other two groups. This is particularly true in attitude toward Capital in Tests A and B.

The position of the three groups in relation to Capital in the three tests shows a very interesting trend. In all tests the rural group is shown to possess the most unfavorable attitude toward Capital while the city group is relatively strong in support of Capital, especially in Tests A and C. These relationships are presented in Figure 1. It should be emphasized that the significance of the relationships existing among these various groups which have just been discussed does not lie in the relative magnitude of the scores but in their trend. In only two cases were differences found which were as much as four times their probable error. The only claim made in connection with the above presentation is that certain relatively small differences show fairly definite trends.

The mean scores on Labor do not show important trends. In Test A the only difference of much importance is that between the city and rural groups. Here, the latter shows a more favorable attitude toward Labor than the former. This, however, is offset by the opposite tendency in Test B. Again, the city group shows a more favorable attitude toward Labor in Test B than the entire group, and is especially more favorable than the town group. This is offset by opposite but weaker tendencies in Test A. The most that may be said of the attitude of all groups toward Labor is that it is favorable. Since the differences are of relatively small magnitude, none being as much as four

TABLE I
MEAN SCORES, STANDARD DEVIATIONS, AND PROBABLE ERRORS OF NET SCORES FOR ALL GROUPS

Group	Test A			Test B			Test C		
	Mean	Standard Deviation	Probable Error of Mean	Mean	Standard Deviation	Probable Error of Mean	Mean	Standard Deviation	Probable Error of Mean
Government									
City	5.52	8.76	.224	4.37	7.82	.37			
Town	5.49	7.98	.378	4.72	7.92	.38			
Rural	6.70	8.73	.410	4.18	7.92	.37			
Combined	6.02	8.20	.224	4.41	7.89	.22			
Capital									
City	0.49	6.92	.328	-0.50	5.74	.27	2.49	7.50	.36
Town	-0.76	5.92	.280	-0.24	5.76	.27	2.19	7.56	.36
Rural	-1.66	6.45	.310	-1.05	5.38	.25	0.28	7.86	.37
Combined	-0.74	6.56	.179	-0.70	5.54	.15	1.64	7.68	.21
Labor									
City	4.50	6.30	.298	2.53	4.04	.191			
Town	5.16	7.20	.340	1.69	4.54	.215			
Rural	5.37	6.10	.347	1.59	4.10	.192			
Combined	4.96	6.84	.186	1.94	4.35	.118			

TABLE II
OBTAINED DIFFERENCES IN MEAN SCORES, PROBABLE ERRORS OF DIFFERENCE, AND CHANCES OF TRUE DIFFERENCE BEING GREATER THAN ZERO BY GEOGRAPHICAL GROUPS

Group	Test A			Test B			Text C		
	Mean Difference	Probable Error of Difference	Chances in 100 That True Differences Greater Than Zero	Mean Difference	Probable Error of Difference	Chances in 100 That True Differences Greater Than Zero	Mean Difference	Probable Error of Difference	Chances in 100 That True Differences Greater Than Zero
Government									
Combined and City	.50	.47	76.0	.07	.43	50.0			
Combined and Town	.53	.45	78.5	.31	.44	68.0			
Combined and Rural	.68	.47	84.0	.23	.43	63.5			
City and Town	.03	.56	50.0	.38	.53	68.0			
City and Rural	1.18	.58	91.0	.16	.52	58.0			
Town and Rural	1.21	.56	92.5	.54	.53	75.0			
Capital									
Combined and City	-1.23	.37	98.7	.20	.30	67.0	.85	.41	91.5
Combined and Town	.02	.33	50.0	.46	.30	84.0	.55	.41	82.0
Combined and Rural	.92	.36	95.5	.35	.28	76.0	1.36	.42	98.5
City and Town	1.25	.44	97.0	.26	.37	68.0	.30	.51	66.0
City and Rural	2.15	.45	99.9	.55	.36	94.5	2.21	.52	99.9
Town and Rural	.90	.42	92.5	.81	.36	93.5	1.91	.52	99.0
Labor									
Combined and City	.46	.35	81.0	.59	.23	96.0			
Combined and Town	.21	.39	63.0	.25	.25	75.0			
Combined and Rural	.41	.40	78.0	.35	.25	83.0			
City and Town	.66	.45	80.0	.84	.29	97.7			
City and Rural	.87	.46	90.0	.94	.27	99.0			
Town and Rural	.21	.49	61.0	.10	.29	51.0			

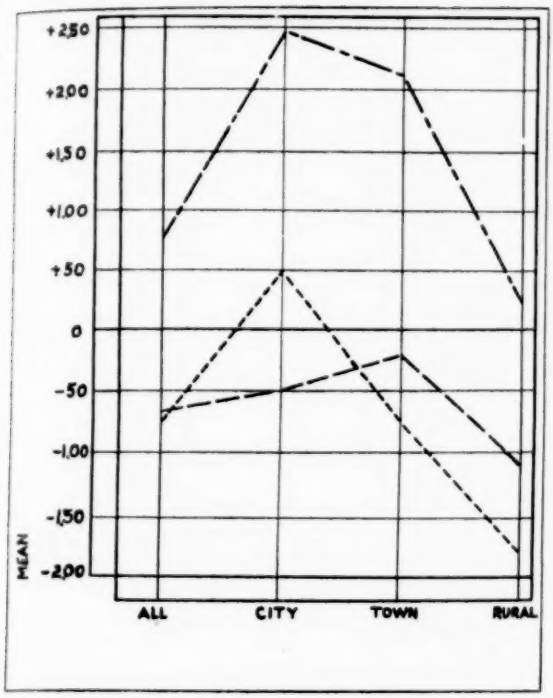


Figure 1. Trend of mean scores, comparing mean scores of entire group with each geographical division in Capital, Tests A, B, and C.
– – – – – – Test A
— — — — Test B
— · — · — Test C

times its probable error, and no uniform trends are shown, little significance can be attached to the existing differences.

Summarizing the differences in attitude toward Government, it may be said that the chief difference is found in the mean of the rural group in Test A. This group is more favorable toward the government than either of the others. In Capital, the two important items are the relatively strong tendency of the city group to support our economic order, and the pronounced drop in the curve indicating relatively strong opposition to Capital on the part of the rural group. The entire group gives relatively strong support to Labor, but the scores show no trends from group to group.

Grade Differences

There are certain grade differences and grade trends in attitude that deserve careful consideration. The first item to which attention should be called is the appearance of a regular grade trend in attitude toward Government both in Tests A and B. These data are to be found in Tables III and IV. The relationships are presented graphically in Figure 2. It will be noted from a consideration of the graph that the curve shows a remarkable rise at the eighth grade. The curve is especially high in Test A, rising to a mean of 8.21 which is more than twice that of the seventh grade.

An investigation was made to discover whether this high mean for the eighth grade was due to an unusually high mean for some particular group. Such was not the case. In all groups in Test A and in all but the city group in Test B, the eighth grade stands high above both the seventh and ninth grades. In order that this may be made clear, the mean for each of these three grades for Test A is presented below:

	Grades		
Group	7	8	9
City	3.38	7.11	5.31
Town	3.07	8.25	6.00
Rural	6.54	9.60	3.38

This difference in the eighth grade mean becomes more remarkable because of the relatively wide geographical areas from which these scores were obtained. The city and town groups were taken from points approximately 130 miles apart, while the rural scores were obtained from six different schools scattered over a fifty-mile area.

As a further check against these scores, the means of another group, which will be discussed later under the title of the Rugg group, are presented for comparison. The scores of this group come from two widely separated towns each with a population of approximately 4,000. The scores for this group are: seventh grade, 3.64; eighth grade, 6.18; and ninth grade, 0.72. It appears, therefore, that this difference in the eighth grade mean cannot be attributed to an error in sampling.

TABLE III
MEAN SCORES, STANDARD DEVIATIONS, AND PROBABLE ERROR OF MEAN BY GRADES

Grade	Test A			Test B			Test C		
	Mean	Standard Deviation	Probable Error of Mean	Mean	Standard Deviation	Probable Error of Mean	Mean	Standard Deviation	Probable Error of Mean
Government									
7	4.08	9.06	.55	2.95	8.40	.51			
8	8.21	8.34	.65	5.46	8.04	.55			
9	4.90	9.18	.60	3.47	8.13	.53			
10	6.48	7.32	.48	4.11	7.50	.49			
11	7.71	7.56	.57	5.76	6.42	.48			
12	5.57	8.13	.54	5.39	7.98	.54			
Capital									
7	-1.46	6.58	.40	-1.30	5.94	.36	0.78	6.78	.42
8	.73	6.52	.45	-0.84	5.08	.35	2.43	7.56	.52
9	-1.15	6.64	.44	-1.80	5.50	.36	0.36	8.10	.54
10	-1.12	6.48	.42	-0.04	5.24	.34	2.12	7.50	.49
11	-0.63	6.28	.47	0.13	5.38	.40	3.35	8.46	.63
12	-2.66	6.26	.42	-0.66	6.00	.40	2.12	8.37	.57
Labor									
7	4.75	7.08	.43	1.00	4.36	.27			
8	4.22	7.74	.53	1.95	4.54	.31			
9	3.53	6.45	.42	1.99	4.40	.29			
10	5.46	7.08	.46	2.11	4.10	.27			
11	4.57	8.28	.62	2.37	4.40	.33			
12	5.72	6.45	.43	2.19	4.28	.29			

Other differences worthy of note are in connection with the eleventh grade. This grade stands higher in its Government score than any other except the eighth. In Test B the Government score of this grade stands out with a mean somewhat higher than that of the eighth grade. The interrelation of differences in Government scores, Test B, shows fairly significant differences in most cases, with a trend toward higher scores from the seventh to twelfth grade.

The mean scores in Capital show important differences, but fail to indicate the definite grade trends that were present in the Government scores. In Test A, the eighth grade maintained the same high place with relation to the other grades, as in Government, being the only grade that showed an appreciable plus score. The tendency of the twelfth grade to deviate from the other grades by showing a somewhat weaker support of Capital is similar to that found in Government. The difference is quite marked, especially in Test A. These relationships are presented in graphic form in Figure 3. Test C in Capital shows relationships similar to those found in Test A. The eighth grade stands high but not as high as the eleventh, while the downward trend of the curve from the eleventh to twelfth grades, which has already been noted in the two preceding tests, is present here. Wide differences are also shown between grades seven and eleven and between nine and eleven.

Significance of Differences in Mean Scores

The discovery and presentation of differences in the mean scores of the various groups under consideration is a comparatively simple matter, but a discussion of their significance is a difficult problem. An attempt is made in the following discussion to

TABLE IV
OBTAINED DIFFERENCES IN MEAN SCORES, PROBABLE ERRORS OF DIFFERENCE, AND CHANCES OF TRUE DIFFERENCE BEING GREATER THAN ZERO BY GRADES

Grade	Test A			Test B			Test C		
	Mean Difference	Probable Error of Difference	Chances in 100 That True Differences Greater Than Zero	Mean Difference	Probable Error of Difference	Chances in 100 That True Differences Greater Than Zero	Mean Difference	Probable Error of Difference	Chances in 100 That True Differences Greater Than Zero
Government									
7 and 8	4.13	.85	99.9	2.51	.75	99.0			
7 and 9	.82	.81	75.0	.52	.73	68.0			
7 and 10	2.40	.73	99.0	1.16	.77	84.0			
7 and 11	3.63	.79	99.9	2.81	.70	99.9			
7 and 12	1.49	.77	91.0	2.44	.74	99.0			
8 and 9	3.31	.88	99.0	1.99	.75	95.5			
8 and 10	1.73	.81	92.5	1.35	.73	89.0			
8 and 11	.50	.86	65.0	.30	.73	61.0			
8 and 12	2.64	.84	98.0	.07	.77	50.0			
9 and 10	1.58	.83	89.0	.64	.72	73.0			
9 and 11	2.81	.82	99.0	2.29	.71	99.0			
9 and 12	.67	.80	72.0	1.92	.76	95.5			
10 and 11	1.23	.74	89.0	1.65	.69	95.0			
10 and 12	.91	.72	81.0	1.28	.73	88.0			
11 and 12	2.14	.78	97.0	.37	.72	63.0			
Capital									
7 and 8	2.19	.60	99.0	.66	.66	81.0	1.56	.67	94.0
7 and 9	.31	.59	63.0	.50	.51	75.0	.42	.69	66.0
7 and 10	.34	.58	65.0	1.26	.50	95.0	1.34	.65	91.5
7 and 11	.83	.62	81.0	1.43	.54	96.5	2.57	.76	99.0
7 and 12	1.20	.58	92.0	.64	.54	78.5	1.34	.71	82.0
8 and 9	1.88	.62	98.0	.96	.50	90.0	1.98	.75	96.5
8 and 10	1.85	.62	98.0	.80	.49	86.5	.22	.71	58.0
8 and 11	1.36	.65	92.0	.97	.53	89.0	1.01	.82	79.0
8 and 12	3.39	.62	99.9	.18	.53	59.0	.22	.77	57.5
9 and 10	.03	.61	50.0	1.76	.50	99.0	1.76	.73	95.0
9 and 11	.52	.64	71.0	1.93	.54	99.0	2.99	.83	99.0
9 and 12	1.51	.61	95.0	1.14	.54	92.0	1.76	.78	93.5
10 and 11	.49	.63	70.0	.17	.53	58.5	1.23	.80	84.5
10 and 12	1.44	.60	95.0	.62	.53	78.5	0.00	.75	50.0
11 and 12	2.03	.63	98.0	.79	.57	82.5	1.23	.85	84.0
Labor									
7 and 8	.53	.68	69.5	.95	.41	94.0			
7 and 9	1.22	.60	91.5	.99	.39	95.0			
7 and 10	.71	.66	78.0	1.11	.37	98.0			
7 and 11	.18	.75	57.0	1.73	.42	99.9			
7 and 12	.97	.60	86.0	1.19	.39	98.0			
8 and 9	.69	.68	75.0	.04	.42	50.0			
8 and 10	1.24	.70	88.0	.16	.41	60.0			
8 and 11	.35	.81	61.5	.78	.44	89.0			
8 and 12	1.50	.68	93.0	.24	.42	65.0			
9 and 10	1.93	.62	98.0	.12	.39	58.5			
9 and 11	1.04	.75	82.5	.74	.44	87.0			
9 and 12	2.19	.60	99.0	.20	.40	63.0			
10 and 11	.89	.77	78.0	.62	.42	84.0			
10 and 12	.26	.62	61.0	.08	.39	50.0			
11 and 12	1.15	.75	84.5	.54	.44	79.0			

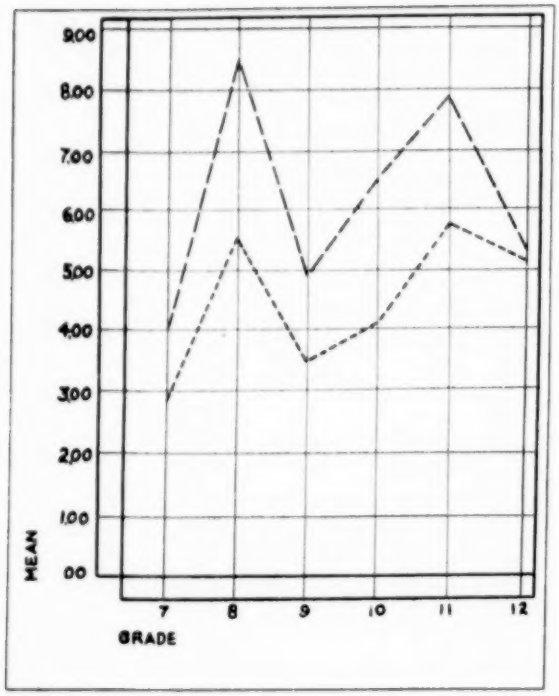


Figure 2. Comparison of grade trends of mean scores in Government, Tests A and B.
———— Test A
------ Test B

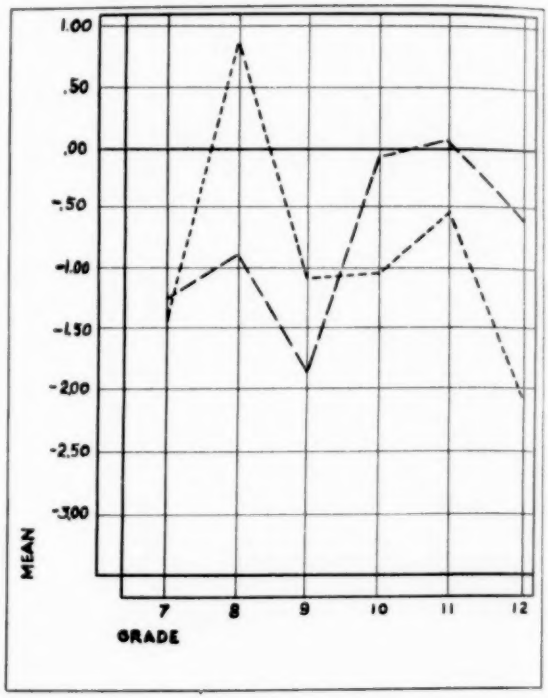


Figure 3. Comparison of grade trends of mean scores in Capital, Tests A and B.
----- Test A
— — — Test B

point out the significance of the general trends of scores and to assign reasonable causes for some of the more important differences which appear. No explanation of any phenomenon can be complete without an adequate knowledge of the important facts with which it is associated. The obvious difficulty in attempting an explanation of the various differences exhibited in these tests lies in the fact that there may be important factors influencing the outcomes which the tests do not reveal. For example, one does not know to what extent the attitude of any particular grade has been influenced by the attitude and opinions of the classroom teacher. One does not know to what extent factors lying entirely outside the classroom may have influenced attitudes.

Thurstone (14) has shown by recent experiments that the attitudes of high school pupils are measurably affected by moving pictures. Again, we cannot be certain of the extent to which children's attitudes are colored by the heat of a political campaign being waged in the country. These and many other factors which cannot be discovered and evaluated in a single study may exert influences upon children's attitudes toward political and economic questions.

The first item to be discussed is the fact that all groups showed a rather pronounced support of our government. This fact when taken by itself should cause no surprise. It would perhaps cause grave concern, even to those who are outspoken critics of our government, to find that the majority of school children were unwilling to support the legally constituted government. The situation appears in a somewhat different light when attention is directed to the type of statements which the majority of these pupils accepted in registering this support. Each statement is supposed to tap a field of controversy in which relatively large numbers

of people disagree. The significance of the outcome lies in the fact that the mean scores were very high in the light of the foregoing facts. It suggests that indoctrination has reached a relatively high point in the teaching of American history and government.

Quite in contrast to the scores in Government are those recorded in Capital. Here is found a pronounced tendency toward attitudes which are unfavorable to our economic system. The question arises at once whether this is not inconsistent with the attitude expressed toward the government. It is obvious to the well-informed adult that our economic system is inextricably intertwined with our political system, and that the former depends for its existence upon the support of the latter. How then can we explain the two opposite tendencies on the part of these pupils?

A partially satisfactory answer to this question may be found in the fact that the statements in the tests pertaining to Government bring into question the integrity of "My Country." The religion of nationalism has been well taught in our schools. Any statement which questions the motives of America and those who direct her destiny is likely to strike fire in the mind of the school child. His history books have been well censored. Evidence is not lacking to show the solicitude evidenced by various patriotic organizations and political groups in protecting our schools against historical heresy (8, 9). School children have learned their lesson well. Are elections in the United States won by money? "Certainly not" is the orthodox answer of more than 86 per cent of the group. Was the territory of the United States gained honestly? "Certainly" comes the answer of three fourths of the group.

We do not seem to have been as successful, however, in teaching the support of our economic system. Organized control of economic teaching in the public schools is of relatively recent origin and consequently has not yet reached that high state of organization which is exhibited in the control of the teaching of history and government. Neither to the teacher nor to the pupils does "My Capitalism" possess the same connotation as "My Country." They cannot be sung to the same tune. In other words, the pattern responses to questions of economic import have not yet become household words; consequently, they are less colored by indoctrinated beliefs. It may not be too fantastic even to suggest that for this reason the responses to the statements in Capital may express a higher degree of thoughtfulness because of greater intellectual freedom.

The political conservatism of the rural group has been noted and may deserve a word of comment. There is a pronounced belief in some quarters that our farm people are becoming radical. Occurrences of recent date in certain farm areas in the middle west give some ground for this belief. Whatever may be claimed as to farmer attitude in general, there is strong reason for believing that the scores resulting from this test reflect the political attitude of the area in which the test was given. The basis for this belief lies in the fact that the area in question is in a congressional district which has always elected a Republican congressman. Even during the Democratic landslide of 1932 this district was safely Republican.

The scores in Capital among the geographical groups in question have been discussed at length in the preceding pages. It remains only to comment briefly on the fact that the rural group indicates a much stronger opposition to our capitalistic system than any of the other groups. It is possible that this difference, particularly between the rural and city groups, may be accounted for by the stronger influence of capitalistic propaganda in the city than in the country. The capitalistic class is chiefly a city class. It would be natural to expect that its influence would be stronger in the city than in the country. Another factor in the particular areas studied is that these farmers are situated just on the edge of the Chicago dairy region in which the economic relations between milk producers and distributors have been undergoing a severe strain for the past year or two.

It has been noted that a regular grade trend exists indicating a progressively stronger support of the government from grade to grade. In the light of our previous discussion this trend would be expected.

The persistent emphasis on nationalism in the schools, to which attention has been called, together with the usual community condemnation of political radicalism seems to bear its fruit from grade to grade. The peculiar drop in the curve from the eleventh to the twelfth grade, as noted in Figure 2, may be attributed in part to the fact that in this grade a more critical study is made of American history and government. This may be augmented somewhat by that usual attack of "senioritis" in consequence of which this group tends to condemn everything. The downward trend of the curve in Capital, as noted in Figure 3, may be explained upon the same basis. It is here that the study of economics is introduced, and in one of the groups represented in this investigation sociology also is given.

The rather disturbing position of the eighth grade groups seems to be partly explained by the fact that introduction of the study of civics occurs at this time in most groups studied. While it is impossible to present objective evidence that this rather spectacular rise in the curve at the eighth grade has this as its cause, it may reasonably be inferred that such is the case. In the first place, the textbooks used in eighth grade civics were of the orthodox type. The elementary civics texts were Hughes' Elementary Community Civics, Dunn's Community Civics, and Finch's Everyday Civics. It may be noted in this connection that one of the above texts has been declared "good" by the Illinois Committee on Public Utility Information (4). Furthermore, it has been the writer's observation that very little is done by way of a critical study of civics in the eighth grade. The influence both of the textbook and the teacher would tend to instill a somewhat blind faith in our government and its economic organization.

Another factor which possibly has influenced the eighth grade scores is that by the time pupils have advanced well into the eighth grade the nationalistic tone of history texts may begin to make itself felt. Even with a liberal textbook in the hands of pupils, there still is present the influence of the orthodox classroom teacher which has not been measured.

Thus, this first introduction of these pupils to the study of government would perhaps have a strong tendency to produce a rather pronounced result in the direction indicated. It cannot be known to what extent the results have been influenced by the fact that these eighth grade pupils have developed an attitude of responsibility by virtue of being the highest class in the school. In all schools studied the plan of eight elementary grades and a four-year high school course was in operation. It is customary for teachers to impress upon eighth grade pupils their responsibility for the ideals of the school. To what extent this might be carried over into political and economic attitudes is mere conjecture, although it is perhaps worth mentioning as a possible contributing factor in the solution of the problem.

QUALITATIVE ANALYSIS OF DATA

Although it is important to know the mean scores of the various groups on the test as a whole, these scores do not reveal the whole truth. Hence, a qualitative analysis of certain factors is presented here.

High and Low Scores

High scores may be made on any of these tests only by assuming a uniformly extreme position either in a plus or minus direction. A few pupils obtained extremely high scores. In fact, one pupil received the highest possible score to be had in Government on Test A which was 26.4, while four received the highest possible score on the test in Labor which was 24.9. This means that these five pupils had no minus scores. Three pupils received the highest possible score, 12.2, in Labor on Test B. One pupil in the entire group reached the highest possible minus score, which was -17.1 in Test B on Capital.

The extremely low scores indicate a somewhat different situation than that represented by extremely high scores. A low score may be obtained in one of two ways. The pupil may mark a number of extreme plus statements and then offset his high plus score by marking a number of extreme minus statements, or he may assume a relatively

neutral position on all statements. In either case his net score, i.e., the algebraic sum of his gross scores, will be low. It seems improbable, however, that these two types of low net scores would indicate the same attitude, although the numerical values would be the same or nearly so.

This may be illustrated by two examples. Pupil 38, an eighth grade rural girl, received a net score of -1.1 in one test, and pupil 74, a ninth grade rural girl, received a net score of 0.2. These scores would seem to indicate that the attitudes of these two pupils were very similar. But a further investigation reveals that the net score of pupil 38 was composed of gross scores of 11.8 and -12.9, while the net score of pupil 74 was composed of gross scores of 4.6 and -4.4.

This situation suggested a study of the composition of a number of low scores to ascertain whether they were the result of high plus and minus scores balancing each other, or whether they were the result of a uniformly neutral position throughout the test. A low score was arbitrarily considered to be any net score not greater than ±2.5, while a high score was considered to be any net score of not less than ±10.0. In the Capital test the basis of selection was ±2.0 and ±7.0, due to the fact that the range of score on this test is narrower than in the Government or Labor sections.

All low scores on Government and Labor in Test A and on Capital in Test B were investigated, there being a total of 373 such scores. Of this number it was found that only 24 scores, or 6.4 per cent of the total, were composed of the sum of large scores. Thus it seems fairly certain that low scores indicate a tendency toward a neutral position maintained rather consistently throughout the test, rather than toward high plus and minus scores balancing each other.

Extreme Scores

Two investigations were made of extreme scores, one to determine whether extreme scores were characteristic of any particular geographical group, and the other to discover the tendency of the group to assume extreme positions on the various statements. Test A was used for the former investigations and Test B for the latter. A net score of ± 20.0 or more was considered as an extremely high score on Government and Labor and ± 12 or more on Capital. Eighty-six such scores were found in Test A. There was no tendency for any group to predominate in these scores. The distribution was city, 28; town, 26; and rural, 32.

The investigation of extreme scores in Test B brought to light some interesting facts. It will be recalled that this is a degree of truth test in which the pupil is asked to underline numbers from 1 to 5 inclusive, indicating the extent to which he believes the statement to be true or false. Underlining number 1 means that he believes the statement to be certainly or completely true, while the underlining of number 5 means that he believes the statement to be certainly or completely false.

Of the entire twenty-one items in this test, 13.9 per cent of the markings of the 612 pupils were on 1, while 24.4 per cent were on 5. These data are found in Table V.

TABLE V
STATEMENTS IN TEST B MARKED "CERTAINLY TRUE" AND "CERTAINLY FALSE" BY GEOGRAPHICAL DIVISIONS

State-ment	City		Town		Rural		Totals	
	True	False	True	False	True	False	True	False
1	12	50	15	46	8	47	35	143
2	17	74	26	53	26	43	69	170
3	53	23	44	30	49	25	146	78
4	15	68	10	80	9	67	34	215
5	48	44	50	57	50	47	148	148
6	18	58	27	71	30	58	75	187
7	36	23	24	28	22	23	82	74
8	32	103	48	67	47	65	127	235
9	23	72	13	87	16	83	52	242
10	39	44	37	35	46	43	122	122
11	6	66	12	41	11	27	29	134
12	16	56	18	31	20	37	54	124
13	9	109	5	107	10	10	24	326
14	33	39	47	51	38	49	118	139
15	27	75	24	75	27	88	78	238
16	26	29	36	36	24	55	86	120
17	25	35	34	38	33	32	92	105
18	9	49	11	44	18	58	38	151
19	40	37	83	16	64	15	187	68
20	47	11	55	17	46	17	148	45
21	42	31	39	36	38	30	119	97
Total	573	1096	658	1046	632	1019	1863	3161

A few examples are given below showing statements upon which the marking was very pronounced upon one side or the other, and some others where the division of opinion was fairly equal.

1. The laws are more favorable to the rich than to the poor. (Certainly true, 35; certainly false, 143)

9. The government of the United States is controlled by the rich. (Certainly true, 52; certainly false, 242)

11. Labor unions are usually disloyal to the government. (Certainly true, 29; certainly false, 134)

15. A poor man has little opportunity to succeed in this country. (Certainly true, 78; certainly false, 238)

20. The Monroe Doctrine is fair to all countries. (Certainly true, 148; certainly false, 45)

Statements upon which extreme scores were fairly evenly divided were:

5. Under the present economic arrangement the rich are getting richer and the poor are getting poorer. (Certainly true, 148; certainly false, 148)

10. Rich people who are accused of crime are seldom convicted by the courts. (Certainly true, 122; certainly false, 122)

14. The Eighteenth Amendment is one of the chief causes of crime. (Certainly true, 118; certainly false, 139)

The rural group accepted fewer statements as certainly true than either of the other groups, the percentages being city 14.4, town 15.3, and rural 11.9. The rural group was also slightly more cautious in rejecting statements by marking them certainly false, although the difference is too small to be of much significance. The percentages were city 25.2, town 24.5, and rural 23.4.

An interesting picture is presented by separating these extreme scores by school grades. If the assumption is made that there are relatively few statements in the field of political and economic controversy, which may be pronounced as certainly or completely true or certainly or completely false it might be justifiable to entertain the hope that as pupils advance from grade to grade they would become progressively less willing to express such high degrees of certainty. Yet the study fails to justify this hope, as is shown by the percentage of statements distributed by grades:

Grade	Percentage
7	24.7
8	24.9
9	23.2
10	23.1
11	26.9
12	23.8

A somewhat better showing was made in the acceptance of statements as certainly true:

Grade	Percentage
7	17.7
8	15.0
9	17.6
10	12.3
11	13.1
12	11.7

Perhaps this last tabulation deserves some comment. An interesting junior-senior high school grouping appears, with the senior group decidedly lower. The average of the percentages for the two groups is 16.7 and 12.3 respectively (Figure 4).

Group Attitude Toward Various Single Statements in Test A

The discussion of mean scores indicated the attitude of the various groups toward the three fields under investigation. However, it does not show the attitude of these groups toward the separate statements which make up the test.

The following section gives: (1) the response of the entire group toward each statement in Test A, (2) the response of each geographical group toward certain statements where significant difference might be expected to exist, and (3) the reaction of various grades to different statements in the test.

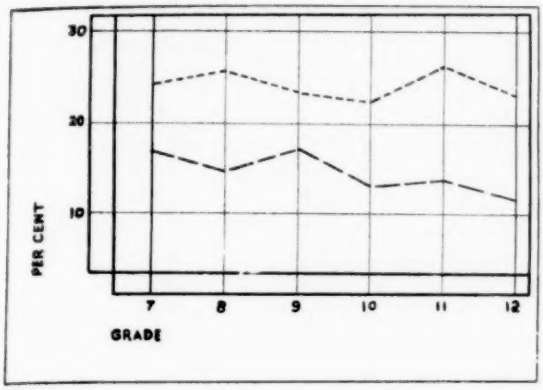


Figure 4. Percentages of pupils marking statements in Test B as "Certainly true" and "Certainly false." Distribution by grades.

----- Certainly false
———Certainly true

The following statements definitely accepted[1] were taken from Test A:

10. Wealthy people are almost exclusively engaged in making money.

17. Laborers should be paid high enough wages so that they may live in comfortable homes.

21. The territory of the United States was gained honestly.

22. American citizens should be given employment in preference to foreigners.

Statements taken from Test A which were definitely rejected are:

2. Criminals in our large cities are shielded by officials.

3. Labor unions are disloyal to the government.

4. Members of labor unions are dangerous persons.

5. Persons who believe in prohibition are dishonest.

8. Persons who talk in favor of Russian communism are patriotic citizens.

12. The laws are more favorable to the rich than to the poor.

13. Elections in the United States are won by the candidate having the most money.

14. Labor agitators are anarchists.

15. Wealthy people should be required to pay a higher rate of taxes than other people.

18. Millionaires came by their wealth honestly.

23. Millionaires have become wealthy by intelligence and hard work.

24. Laborers should be given shorter hours without any reduction in pay.

The following are statements in Test A to which practically neutral responses were given:

1. Elections are controlled by men of wealth.

6. Bankers are interested in making money regardless of the welfare of their depositors.

7. Prices are high because our industries are controlled by corporations.

9. Decisions of the Supreme Court are partial to those who have wealth.

11. Strikes by labor unions ought to be prohibited by law.

16. Public utilities such as electric power plants and telephone systems should be owned by the government.

1. Acceptance or rejection is based on the sum of the total gross score in favor of the statement plus the total gross score in opposition to it. The percentage of gross scores by which the following statements were accepted was from 74.5 to 84.9 per cent. The percentage by which others were rejected was from 61 to 86.8 per cent. In the list in which the response was practically neutral the percentage was approximately 50 per cent for and 50 per cent against.

19. Elections are carried on honestly.

20. Good citizens are in favor of labor unions.

Complete data for these three types of statements are given in Table VI.

Some interesting differences were noticed in the response of the different geographical groups to certain statements. Where a difference in attitude might be expected, the scores of each group were segregated and compared with one another. These differences may best be expressed in graphic form. Figure 5 shows the distribution of the responses to the statement, "Persons who talk in favor of Russian communism are patriotic citizens." This statement was rejected by all groups, but it is significant to note the increasing condemnation of the statement as the curve proceeds from the city to the country group.

Figure 6 indicates the distribution of the response to the statement, "Good citizens are in favor of labor unions." The curve shows a progressive difference again, but in the opposite direction. The majority opinion is favorable to the statement in the city, but unfavorable in both town and country. Mean net scores on each statement by grades for Test A are presented in Table VII. Figure 7 presents the attitude toward the statement, "Laborers should be given shorter hours without any reduction in pay."

Trend of Attitude Shown by Grade Distribution

Some interesting facts may be brought to light by a study of the grade distribution of the attitude toward the various statements in Test A. When a separation of the entire group is made into junior and senior high school divisions, there is evidence that the senior high school group tends more toward a position of conservatism toward Government than the junior high school group. In all but two statements in Test A relating to Government, the senior high school group shows higher mean scores than the junior high school. Although the latter group shows a favorable attitude toward the government in all but one of the statements, its mean score is lower. These data are found in Table VIII.

TABLE VI
TOTAL SCORE OF ENTIRE GROUP IN TEST A, SHOWING PER CENT OF SCORES IN FAVOR AND AGAINST EACH STATEMENT

Statement	Total Score		Per Cent	
	In Favor	Against	In Favor	Against
1	474.8	554.4	46.2	53.8
2	296.4	1114.3	21.7	78.3
3	211.6	1110.9	16.0	84.0
4	181.3	753.0	19.4	80.6
5	194.6	1151.0	14.5	85.5
6	552.7	653.3	45.9	54.1
7	536.3	418.9	56.1	43.9
8	307.5	1166.9	20.9	79.1
9	427.1	411.0	50.9	49.1
10	616.0	180.3	77.3	22.7
11	637.3	588.5	51.9	48.1
12	306.6	1184.4	20.6	79.4
13	159.4	1054.7	13.2	86.8
14	354.2	740.4	32.4	67.6
15	251.1	529.8	32.2	67.8
16	382.0	623.7	38.0	62.0
17	1591.0	281.9	84.9	15.1
18	261.9	947.4	21.7	78.3
19	940.2	578.3	61.9	38.1
20	623.5	608.0	50.5	49.5
21	1032.9	352.2	74.5	25.5
22	941.2	230.1	80.3	19.7
23	173.6	851.9	17.0	83.0
24	571.9	923.7	39.0	61.0

TABLE VII
MEAN NET SCORES ON EACH STATEMENT IN TEST A BY GRADES

Statement	Grades					
	7	8	9	10	11	12
1	-0.235	0.086	0.142	0.036	0.466	0.0
2	1.094	1.799	0.979	1.041	1.817	1.282
3	1.064	1.542	1.262	1.365	1.844	1.500
4	1.240	1.085	0.925	0.681	0.638	0.933
5	1.200	1.592	1.763	1.574	2.385	1.222
6	-0.267	0.329	0.139	0.075	0.469	0.302
7	-0.803	-0.033	-0.235	-0.051	-0.204	-0.211
8	1.101	1.239	1.572	1.357	1.646	1.605
9	-0.575	0.130	-0.260	-0.120	0.437	0.231
10	-0.881	-0.536	-0.591	-0.671	-0.817	-0.760
11	0.072	-0.160	0.016	-0.099	-0.179	0.0
12	1.226	1.949	1.133	1.522	1.595	1.283
13	1.409	1.980	1.733	1.050	1.672	1.485
14	0.841	1.277	0.533	0.843	0.114	0.354
15	0.756	0.927	0.210	0.281	0.462	0.069
16	0.461	0.685	0.195	0.320	0.350	0.548
17	1.794	1.663	2.260	2.413	2.913	1.420
18	-0.913	-1.044	-1.405	-1.206	-1.228	-1.038
19	0.601	1.107	0.685	0.706	0.144	0.465
20	0.341	0.135	-0.026	-0.080	-0.239	-0.034
21	0.790	0.815	0.636	1.505	1.750	1.362
22	0.970	0.380	1.152	1.462	1.591	1.617
23	-1.165	-0.907	-1.289	-1.036	-1.077	-1.143
24	-0.786	-0.975	-0.963	-0.223	-0.137	-0.158

TABLE VIII
TREND OF ATTITUDE SHOWN BY JUNIOR-SENIOR HIGH SCHOOL DISTRIBUTION

Statement	Group	Junior High School (N = 324)	Senior High School (N = 288)
1	Government	0.010	0.141
2	Government	1.261	1.343
3	Labor	1.209	1.547
4	Labor	1.090	0.757
5	Capital	1.500	1.668
6	Capital	0.046	0.266
7	Capital	-0.388	-0.151
8	Capital	1.296	1.456
9	Government	-0.262	0.160
10	Capital	-0.678	-0.743
11	Labor	-0.015	-0.083
12	Government	1.412	1.459
13	Government	1.630	1.378
14	Labor	0.857	0.467
15	Capital	0.630	0.258
16	Capital	0.445	0.409
17	Labor	1.906	2.198
18	Capital	-1.111	-1.150
19	Government	0.834	0.316
20	Labor	0.160	-0.085
21	Government	0.748	1.521
22	Labor	0.815	1.550
23	Capital	-1.128	-1.085
24	Labor	-0.900	-0.178

As has already been pointed out, all mean scores in Capital hovered around zero. In this comparison the senior high school group was somewhat less favorable to Capital than the junior high school, the means being -0.84 and -0.59 respectively.

The mean of the senior high school group shows a decidedly more favorable attitude toward Labor. The means were 7.99 for the senior high school and 4.25 for the junior high school.

Grade differences in the several separate statements were not found to be particularly significant and were somewhat irregular. Some of these more significant differences are shown graphically in Figures 9 to 18 inclusive. These graphs show the typical trend from the junior to senior high school group. In some cases, notably Figures 10 to 15 inclusive, a regular grade trend is shown. These graphs also indicate the tendency of the group to separate into 7-8-9 and 10-11-12 grade divisions. This is particularly noticeable in Figures 9, 12, 16, and 18.

Although the differences expressed in these graphical representations and those immediately following are not large, in most cases their significance may be emphasized by drawing out the scale upon which the scores are based. Figure 8 illustrates this point. The scores are taken from Figure 16. It may be seen from the above representation that although the differences in mean scores of these groups are small, yet the senior high school group is distributed along the scale between "Most" and "All", while the junior high school group lies between "Many" and "Most."

Attitude in Relation to Content Material Contained in Social Science Textbooks

It is interesting to know something of the extent to which attitude may be influenced by the type of textbook material used by pupils. It seemed worth while to make a

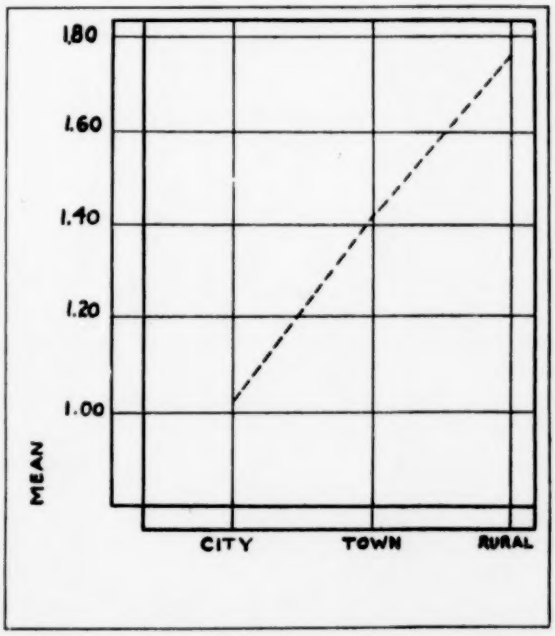


Figure 5. Trend of mean scores in city, town, and rural, on Statement No. 8, Test A "Persons who talk in favor of Russian Communism are patriotic citizens." A plus score tends to reject the statement.
------Test A

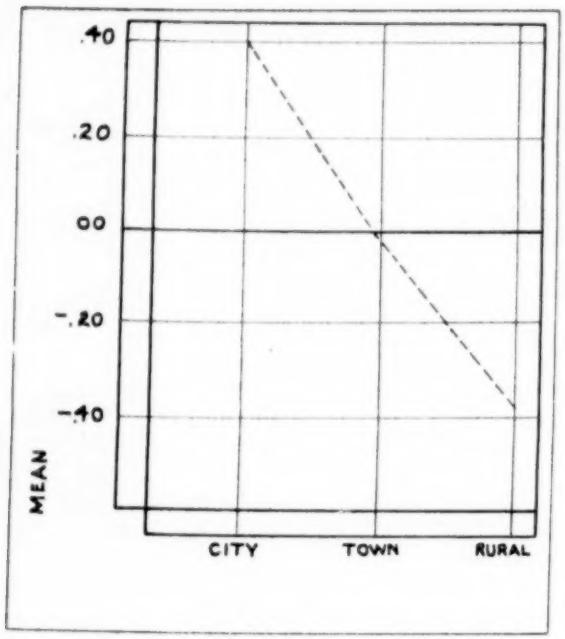


Figure 6. Trend of mean scores in city, town, and rural on Statement No. 20, Test A, "Good citizens are in favor of labor unions." A plus score tends to accept and a minus score to reject the statement.

-------Test A

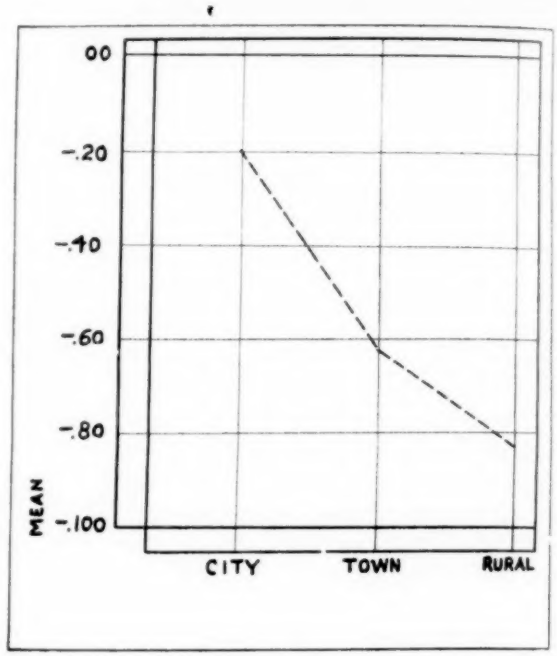


Figure 7. Trend of mean scores in city, town, and rural on Statement No. 24, Test A, "Laborers should be given shorter hours without any reduction in pay." A minus score tends to reject the statement.

------ Test A

brief investigation into this field.

A comparison was made of attitude scores between groups comprising the present study and groups that had used the Rugg social science textbooks. These texts have been used as the basis for the comparison on the assumption that their content, as well as the method of presentation employed in the texts, is distinctly liberal as compared with most other texts in the same field.

The history texts used in the various schools of the regular groups are Tryon and Lingley, The American People and Nation; Halleck, The Story of Our Country; and Leonard and Jacobs, The Nation's History. In civics, the texts used were Hughes, Elementary Community Civics; Broome and Adams, Conduct Through Citizenship; and Finch, Everyday Civics.

Tests A and B were administered in the seventh and eighth grades in a town of 4000 population and in the ninth grade of another

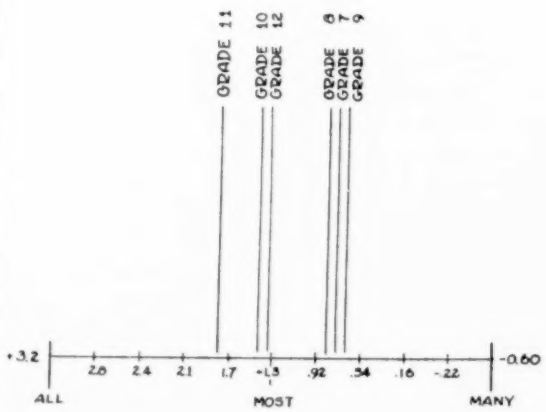


Figure 8. Enlargement of scale showing the distribution of mean scores on the scale by grades. Mean scores taken from Statement 21, Test A, "The territory of the United States was gained honestly."

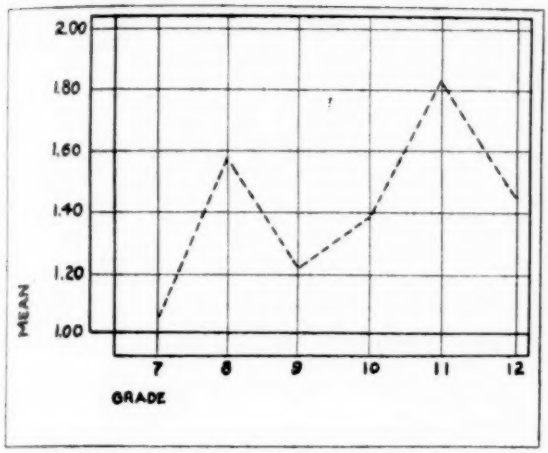


Figure 9. Trend of mean scores by grades on Statement No. 3, Test A, "Labor Unions are disloyal to the government." A plus score tends to reject the statement.

-----Test A

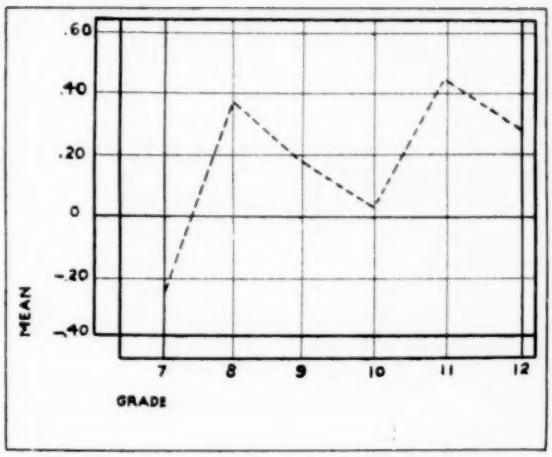


Figure 11. Trend of mean scores by by grades on Statement No. 6, Test A, "Bankers are interested in making money regardless of the welfare of their depositors." A plus score tends to reject and a minus score to accept the statement.

------Test A

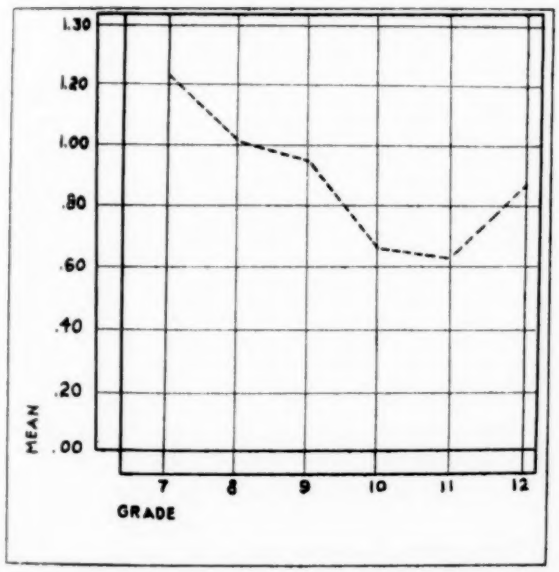


Figure 10. Trend of mean scores by grades on Statement No. 4, Test A, "Members of labor unions are dangerous persons." A plus score tends to reject the statement.

---- Test A

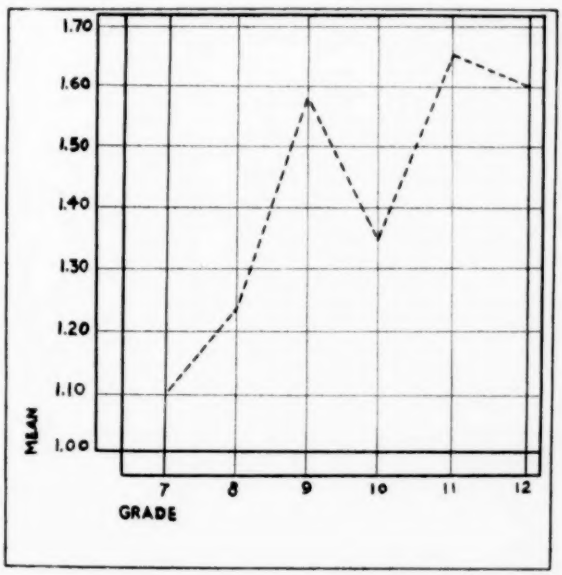


Figure 12. Trend of mean scores by grades on Statement No. 8, Test A, "Persons who talk in favor of Russian Communism are patriotic citizens." A plus score tends to reject the statement.

------ Test A

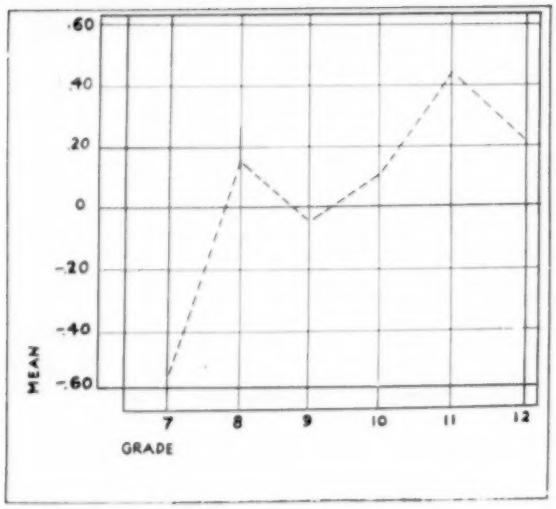


Figure 13. Trend of mean scores by grades on Statement No. 9, Test A, "Decisions of the Supreme Court are partial to those who have wealth." A plus score tends to reject and a minus score to accept the statement.

----- Test A

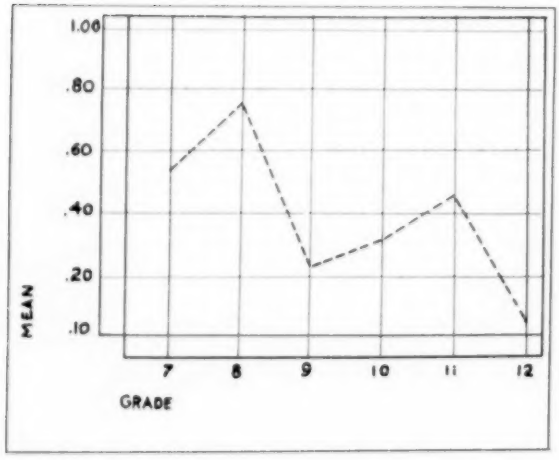


Figure 14. Trend of mean scores by grades on Statement No. 15, Test A, "Wealthy people should be required to pay a higher rate of taxes than other people." A plus score tends to reject the statement.

---- Test A

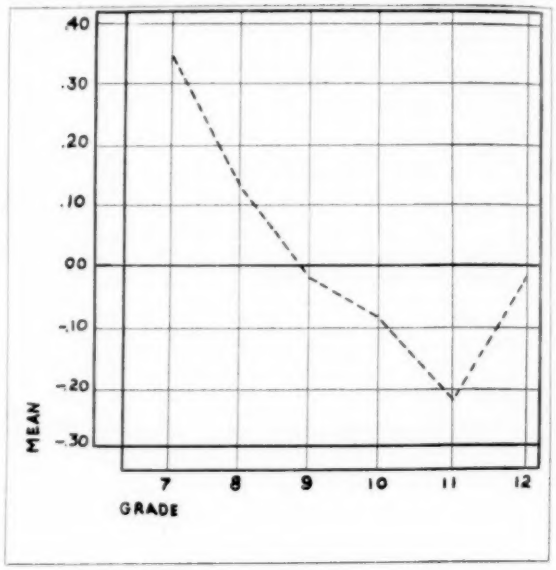


Figure 15. Trend of mean scores by grades on Statement No. 20, Test A, "Good citizens are in favor of labor unions." A plus score tends to accept and a minus score to reject the statement.

-----Test A

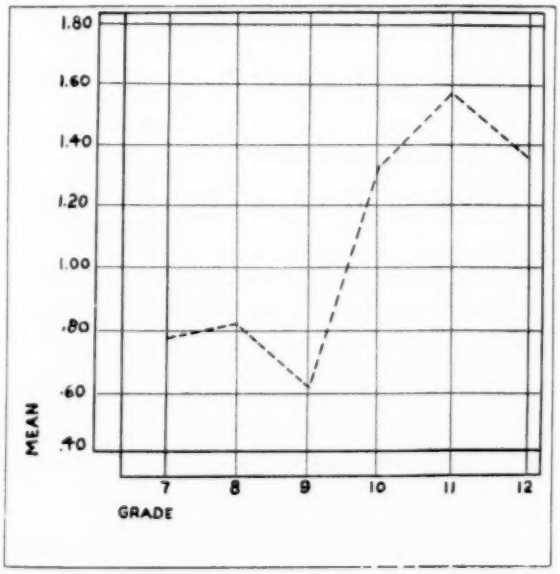


Figure 16. Trend of mean scores by grades on the statement, "Territory comprising the United States was gained honestly." A plus score tends to accept the statement.

---- Test A

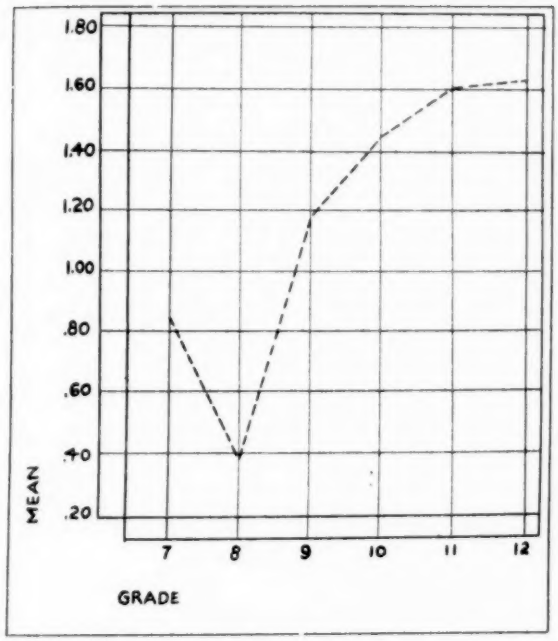


Figure 17. Trend of mean scores by grades on the statement, "Americans should be given employment in preference to foreigners." A plus score tends to accept the statement.

-----Test A.

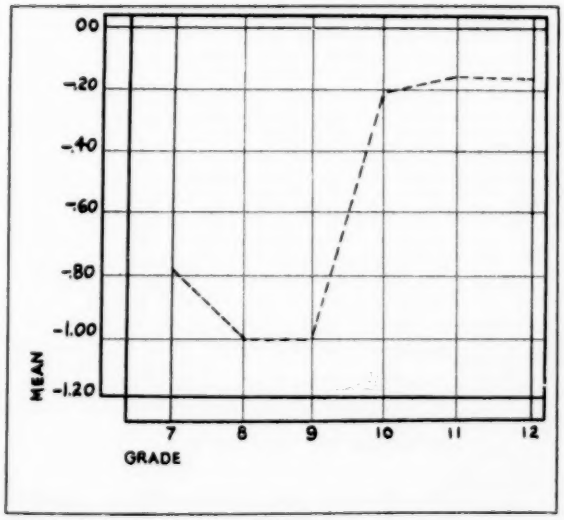


Figure 18. Trend of mean scores by grades on the statement, "Laborers should be given shorter hours without any reduction in pay." A minus score tends to reject the statement.

------ Test A

town of approximately the same size. There were 176 pupils in this group. The mean scores obtained from these pupils were then compared with those of the same grades in the regular group of 612 students. These comparisons revealed some rather significant differences in attitude. The data are presented in Tables IX and X.

TABLE IX
ATTITUDES OF RUGG AND NON-RUGG GROUPS

Group	Mean	Standard Deviation	Probable Error of Mean
	Test A		
	Government		
Rugg	4.45	7.72	.39
Non-Rugg	5.58	9.07	.34
	Capital		
Rugg	-1.24	5.57	.28
Non-Rugg	-0.70	6.68	.25
	Labor		
Rugg	6.93	6.22	.32
Non-Rugg	4.52	7.09	.27
	Test B		
	Government		
Rugg	1.26	4.82	.24
Non-Rugg	3.87	8.28	.31
	Capital		
Rugg	-0.83	5.42	.27
Non-Rugg	-0.78	5.58	.21
	Labor		
Rugg	3.09	3.75	.19
Non-Rugg	1.91	4.38	.16

It will be noted that the Rugg group is distinctly less willing to support our present order of government than the non-Rugg group. The difference in Labor scores is also quite pronounced. In attitude toward Capital, however, the two groups appear quito similar. The relationship between the two groups is presented in Figures 19 and 20.

The number of cases is large enough to indicate a significant difference. No factors other than textbook usage were revealed which might account for these differences. Certainly, neither of the communities from which the Rugg scores were obtained could in

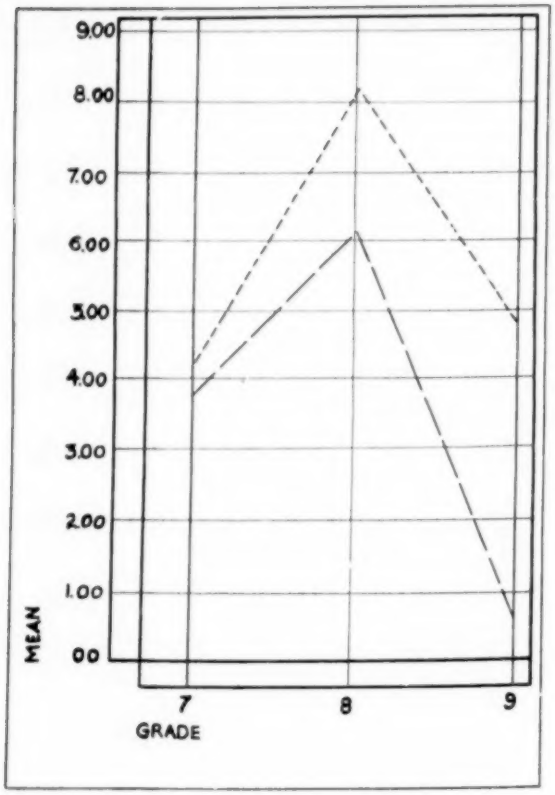


Figure 19. Comparison of mean scores of Rugg and non-Rugg groups, by grades, in Government, Test A.
— — — Rugg
– – – – – Non-Rugg

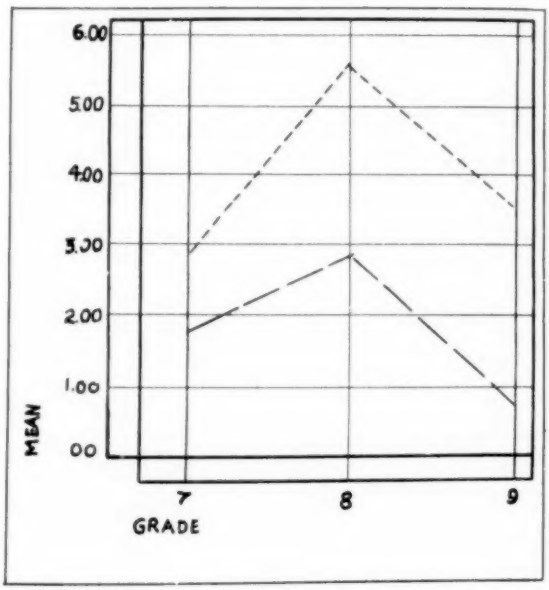


Figure 20. Comparison of mean scores of Rugg, non-Rugg groups by grades in Government, Test B.
— — — Rugg
– – – – – Non-Rugg

TABLE X
COMPARISON OF RUGG WITH NON-RUGG GROUPS SHOWING DIFFERENCES, PROBABLE ERRORS OF DIFFERENCES, AND CHANCES IN 100 THAT TRUE DIFFERENCES ARE GREATER THAN ZERO

Group	Difference	Probable Error of Difference	Chances in 100 That True Differences Greater Than Zero
Test A			
Government	1.13	.52	93.0
Capital	.54	.37	84.0
Labor	2.41	.41	99.9
Test B			
Government	2.61	.40	99.9
Capital	.05	.33	50.0
Labor	1.18	.26	99.9

any sense be classed as radical. On the contrary, the community from which the seventh and eighth grade scores were obtained is reputed to be a highly conservative community. The fact that the mean scores of the Rugg group on both Government tests indicated a more critical attitude toward the government than was shown by the other group is significant, in view of the reputation for liberalism in the Rugg textbooks as contrasted with the rather extreme nationalistic trend of the more commonly used social science textbooks.

The greater similarity of scores on Capital for the two groups under consideration tends to give weight to the explanation of the negative scores for the regular group previously set forth. This explanation was that the school propaganda for economic conservatism has not been carried on as effectively as it has for political nationalism.

Correlation with IQ and Socio-Economic Status

For purposes of this part of the investigation, IQ's based on the Otis Self-Administering Test of Mental Ability were obtained for 188 pupils in the city group. Socio-economic scores were secured from the entire city group of 202 pupils by means of the Sims Score Card. Correlations were computed to determine the relation between intelligence and the attitude of favorableness toward Government, Capital, and Labor.

In computing the correlations, both the minus and plus scores were used in the series on the assumption that all scores lying below the highest plus score have an ingredient of unfavorableness. So long as favorableness outweighs unfavorableness, as measured by the test, the score will be plus. When the two appear in practically equal amounts, the score will be zero. Similarly, when favorableness is outweighed by unfavorableness, the score will be minus; only in the most extreme minus scores is the ingredient of favorableness practically absent.

There was no correlation between intelligence and any one of the three fields. No correlation was found between the attitude of favorableness toward either the Government or Capital and socio-economic status, but a negative correlation of .39 with a P.E. of .07 between the attitude of favorableness toward Labor and socio-economic status was obtained. It might be expected that a correlation of this kind would exist, since those of low socio-economic status in a city will almost invariably belong to the wage-earning class.

Contributions of Study to Civic Training

Public school education has as one of its goals the inculcation of desirable civic attitudes. Educators have heretofore been satisfied to determine the effectiveness of teaching and the progress of pupils chiefly, if not entirely, on the basis of measured results in terms of the content material of the curriculum. This is not surprising in view of the fact that until recently no satisfactory measuring instruments have been available for determining growth in citizenship. The techniques employed in this study offer a further means for measuring the effectiveness of teaching. However, they do not furnish a basis for determining what is desirable in civic attitudes. They do not apply to the goals themselves, but they can be used to evaluate materials, methods, curricula, and teachers.

SUMMARY AND CONCLUSIONS

The purpose of this investigation was to discover and measure the attitudes toward current political and economic problems of pupils in grades seven to twelve inclusive; to discover the extent to which attitudes are influenced by certain factors such as place of abode (whether in city, town, or country), socio-economic status, school grade, and general intelligence; and to suggest some contributions to civic training arising from the study.

An attitude test was prepared by the writer and administered to 612 pupils in six different schools. A subsidiary study was made of 176 pupils located in two different towns to determine the effect of textbook material upon pupil attitudes.

The following are some of the more important results of the study:

1. The mean scores of all groups indicated a distinctly favorable attitude toward the present order of government.
2. The rural group stood measurably higher in its support of Government than either town or city.
3. The tendency was toward a progressively stronger support of Government from grades seven to eleven, but a decline in the mean score of the twelfth grade.
4. The attitude of the group as a whole was slightly unfavorable to Capital in two tests, but favorable in a third test by a relatively small margin.
5. Opposition to Capital grew progressively from city to town and to country.
6. No distinct grade trends in attitude toward Capital were discovered, although the eighth grade showed a tendency toward its support while the twelfth grade indicated a rather pronounced opposition as compared with the two preceding grades.
7. The entire group indicated a pronounced

support of Labor, with no definite trends from group to group.

8. A regular, although not pronounced, grade trend indicated a progressively stronger support of Labor from the seventh to twelfth grades.

9. The study gives evidence of the indoctrination of beliefs in United States history and government. This view is supported by: (1) the nature of the statements to which it was necessary for a pupil to subscribe and to which a majority did subscribe in order to obtain a high plus score in Government, and (2) the fact that the trend from grade to grade was toward a progressively stronger support of Government.

10. There was little evidence that the pupils studied became any more open-minded from grade to grade. Twelfth grade pupils were only slightly less likely to subscribe to extreme statements than those of the seventh grade.

11. There seemed to be a tendency toward attitude grouping on the basis of a junior high school group (grades 7-8-9) and a senior high school group (grades 10-11-12). This was especially noticeable in the test on Government.

12. There is evidence, although not conclusive proof, that textbooks which present political and economic problems from a liberal point of view will influence student attitudes accordingly.

13. The techniques exemplified in this study should make possible a more careful scrutiny of the end products of social science teaching than has heretofore been possible.

REFERENCES

1. Allport, Gordon W. "The Composition of Political Attitudes," American Journal of Sociology, XXXV (1929-30), 220-38.
2. Bain, Read. "An Attitude on Attitude Research," American Journal of Sociology, XXXIII (1927-28), 940-57.
3. Chicago Association for Child Study and Parent Education. "Developing Attitudes in Children," Proceedings of the Mid-West Conference, March, 1932 (Chicago: University of Chicago Press, 1933), p. 156.
4. Gruening, Ernest. The Public Pays: A Study of Power Propaganda (New York: Vanguard Press, 1931), pp. ix + 273.
5. Katz, Daniel, and Allport, Floyd Henry. Students Attitudes: A Report of the Syracuse University Reaction Study (Syracuse, N. Y.: Craftsman Press, 1931), pp. xxviii + 408.
6. Murphy, Gardner, and Murphy, Lois Barclay. Experimental Social Psychology (New York: Harpers, 1931), pp. ix + 709.
7. Neumann, George Bradford. A Study of International Attitudes of High School Students: With Special Reference to Those Nearing Completion of Their High School Courses. Teachers College Contributions to Education, No. 239 (New York: Teachers College, Columbia University, 1926), pp. vi + 120.
8. Pierce, Bessie Louise. Civic Attitudes in American School Textbooks (Chicago: University of Chicago Press, 1930), pp. xvi + 297.
9. Pierce, Bessie Louise. Citizen's Organizations and the Civic Training of Youth (New York: Scribner's, 1933), pp. xvii + 428.
10. Porter, E. "Student Opinion on War." University of Chicago, Unpublished Doctor's Dissertation, 1926. (Quoted by Murphy, Gardner, and Murphy, Lois Barclay. Experimental Social Psychology. New York: Harpers, 1931, pp. ix + 709.)
11. Rugg, Harold. An Introduction to American Civilization: A Study of Economic Life in the United States: A Textbook in Geography and Civics with Historical Backgrounds (Boston: Ginn and Company, 1929), pp. xv + 610.
12. Rugg, Harold. A History of American Civilization (New York: Ginn and Company, 1930), pp. xvii + 636.
13. Rugg, Harold, A History of American Government and Culture: America's March toward Democracy (Boston: Ginn and Company, 1931), pp. xvii + 635.
14. Thurstone, L. L., and Chave, E. J. The Measurement of Attitude: A Psychophysical Method and Some Experiments with a Scale for Measuring Attitude toward the Church (Chicago: University of Chicago Press, 1929), pp. xii + 96.
15. Watson, Goodwin B. The Measurement of Fair-mindedness. Teachers College Contributions to Education, No. 176 (New York: Teachers College, Columbia University, 1925), p. 97.

APPENDIX

INSTRUCTIONS TO JUDGES

Instructions for Sorting Statements of Opinion on Certain Economic and Political Questions

Tests A and B

You are being handed herewith a packet containing statements expressing certain attitudes on the following subjects:

(G) The established order of government in the United States

(C) Capitalism as embodied in our American economic practice

(L) Labor (including both organized and unorganized labor)

You are asked first, to sort these statements into four piles, pile No. 1 to contain those statements which refer primarily or exclusively to government, No. 2 those which refer at least principally to capitalism (our economic order), No. 3 those which refer to labor. There may be a few statements which in your judgment do not relate to any of the subjects named above. If so, place such statements in a separate pile, place a band around them and label this pile X.

Now begin with your pile of statements relating to government and distribute them as follows: Visualize a scale before you containing eleven steps thus:

Most unfavorable	Scale											Most favorable
	1	2	3	4	5	6	7	8	9	10	11	

Place on step 1 of the scale all statements expressing an attitude that is completely unfavorable toward the established order of government in the United States and on step 11 all statements expressing the most extremely favorable attitude toward the established form of government. Place on step 6 those statements which in your judgment show an attitude which is neutral or practically so.

Distribute the remaining slips upon the scale so as to express the varying degrees of favorableness or unfavorableness toward the present order of government.

You may not need to use all of the eleven steps on the scale in distributing the statements but use them all if you think it necessary. If you do not find any statements that indicate the most extremely favorable or unfavorable attitudes, you will not need to use steps eleven or one, but only nine or ten and two or three as the outside steps.

Now number the top slip of each pile so as to indicate the step on the scale which each pile represents. Your first pile then would be No. 1 and so on through the scale, the last pile being No. 11. (This numbering process is very important. Be sure to make no mistake at this point.)

Now place one pile on top of another and place a band around the entire pile and label it G for government.

Follow exactly the same procedure with each of the other piles containing statements referring to capitalism and labor. Mark the pile of slips representing capitalism C and the one representing labor L.

Now place a band around the entire lot including the X pile and return them to your instructor.

Test C

I am making a survey of opinion on economic and political problems among junior and senior high school pupils. In order that the test scores may be as valid as possible I am having responses to the test evaluated by a group of competent judges of which you have been chosen as one.

The test covers attitude in the following fields:

The established order of government in the United States

Capitalism as embodied in our American economic practice
Labor (including both organized and unorganized labor)

Instructions for Marking

1. Read the pupil's instructions on page 2, noting that a pupil marks the statement (+) if he accepts and (-) if he rejects the statement.

2. Beginning on page 3 and continuing to the end of the test are a number of paragraphs under the title, Statement of Fact, each one designated by a Roman numeral. Beneath each Statement of Fact are four responses, some of which are marked (+) by the pupil indicating that he accepts the statement; some are marked (-) indicating that he rejects the statement.

3. It is these responses numbered 1, 2, 3, 4 under each Statement of Fact which you are to evaluate, taking into account in your evaluation whether the pupil has accepted it (+) or rejected it (-).

4. Read the Statement of Fact.

5. Read each of the responses below the Statement of Fact.

6. On the extreme left-hand margin, outside of the parentheses, indicate the field of attitude to which in your judgment each response relates, and whether it is in favor of or against a particular field of attitude. Note: Be sure to understand that it is the pupil's marking (+) or (-) of these responses which you are to evaluate; i.e., what attitude, if any, is indicated by a (+) marking and what attitude by a (-) marking on the part of the pupil.

7. Use the following marking system: C = capitalism, G = government, L = labor group. Plus = in favor of, Minus = against, and Zero = no attitude. Thus, if you believe a response indicates an attitude in favor of capitalism you will mark C on the margin opposite this response, if against capitalism -C, etc.

8. If the response indicates an attitude in more than one field indicate this as follows: L, -G, for example, would mean that the statement indicates an attitude in favor of the labor group and against government.

9. If in your judgment a response indicates a neutral attitude mark the item with a O.

10. Be sure to mark every response.

A SURVEY OF OPINION ON ECONOMIC AND POLITICAL PROBLEMS

Instructions for Administering Test

1. Time allotted for test, 30 minutes, not including time spent in filling in the blanks on the front page of test.

2. After each pupil has been provided with a copy of the test instruct pupils to fill in all blanks on the front cover of the test.

3. Read the following statement to the pupils:

> "In the folder before you are three tests, Test A, Test B, and Test C. You are asked to read the instructions carefully for each test. An illustration is given, showing how to mark each test. Study these illustrations carefully, especially if you are not sure that you understand the instructions.
>
> "Your answers will be looked upon as possessing real value, although the results of the test will in no way affect your school grades. Mark each item thoughtfully and carefully.
>
> "You should be sure to note that a test does not end at the bottom of the page. Watch the top of the next page for the remainder of the test.
>
> "Be sure not to omit any item.
>
> "You will be allowed 30 minutes to complete your work on the three tests.
>
> "Now turn the page to Test A and begin. Continue through the three tests without stopping."

4. In order that the data shall be as valid as possible, teachers are asked that reasonable precautions be taken against pupils' copying their work from one another.